542		- Di	STIC STIC	ight	O.S. DET AIT	AND TRADEMARK OFFI	CE
	į	ONLINE SEAR	CH REQUES	T'FORM	* * * * * *	*****	*
******	******** Socoto	********	SERIAL NUMBER	223	263		
USER	1812	PHONE 308-	-1793	DATE 4	10/95		
Please give	a detailed state	ement of requirent	nents. Describe t may have spec	J.G. 1110-1111-15	as possible t iive examples	the subject or relevant	
You may in	clude a copy of th	ne broadest and or	r relevant claims). '			
	Please	ne broadest and or	.ch. A	¥ 2			
•					4		3
			,			J. See to	
							30
rê.							100000
	•						
							1.6
				A CONTRACTOR OF A SECTION OF A			
A STATE OF		الله الله الله الله الله الله الله الله	n and the second		-700		
n na natural de la companya de la c							\ \.\.\.\.\.\.\.\.\.\.\.\.\.\.\.\.\.\.\
				٠.			
	·	•					
•							
				V	*****	*****	***
*****	*****	* * * * * * * *	* * * * * * * * * * STAFF USE ONLY	* * * * * * * * * * * * * * * * * * *		4-70	-
COMPLET		11 My X4258			YSTEMS CAS ONLINE DARC/QUES DIALOG	TEL	
SEARCHEI ONLINE TII	ME 1/2 TOT	ALTIME		_ _ _	SDC TG		

•• ;

THE STREET STREET

THE PROPERTY OF THE PARTY OF TH

Listing for Mary Hale Tue Apr 11 13:27:35 1995

iGenetics

Fast Pairwise Comparison of Sequences

Results file sqlasq.res made by on Tue 11 Apr 95 9:45:54-PDT.

Query sequence being compared:US-08-223-263-1 (1-353)
Number of sequences searched: 50375
Number of scores above cutoff: 4542

Results of the initial comparison of US-08-223-263-1 (1-353) with: Data bank : A-GeneSeq 17, all entries

R - 0 - F10000- - S - -100000-N -U50000-M -B -E -R -

500-

50-1100-1-1-1-1

Listing for Mary Hale

PARAMETERS

SCORE 01 STDEV -1

07=

_5=

3 15 -

4 12

515

17| 6

20 7

22

	Initial scores to save Optimized scores to save	Similarity matrix Mismatch penalty Gap penalty Gap size penalty Cutoff score Randomization group
250	e 45	Unitary 1 1.00 0.05 2
CEARCH STATISTICS	Alignments to save Display context	K-tuple Joining penalty Window size
	15 10	20 32

Cut-off rai Cut-off rai Cut-off rai Cut-off rai Cut-off rai Cut-off rai	Number of a Number of a Number of a	Times:	Scores:	
raised to 3. raised to 4. raised to 5. raised to 6. raised to 7. raised to 8. raised to 9.	residues: sequences searched: scores above cutoff:	CPU 00:00:50.06	Mean 4	
	6065180 50375 4542		Median 5	
		Total Elapsed 00:00:51.00	Standard Deviation 2.39	•

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Tue Apr 11 13:27:35 1995

Page

ω

45. R37486											P81		R123		30. R04032										20. R07353										10. R12538	R1485	R28	7. R45155		6. R20229	•	•	•
Jojoba fatty acyl reductase e	ant nematopotetto mo	hematopoietic	Bovine adrenocorticotropic ho	and heavy chair	Homologous to chicken nov gen	н	Human recombinant erythropoie	O	human erythropoi	hemato	nce of p	ica seed acyl	appa) chain variable	4 standard deviations	gth T4 encoded by pl	encoded by the 2nd	ia rickettsii p120	encoded by I	portion of E	gondii P66 anti	rotein f	polysaccharide	growth hormone	hormone.		horn	DNA o	ດ	Human Cytomegalovirus antigen	MPIV env protein with growth	deduced from human	Ħ	receptor from	nal mucin deduced	87 intestinal mucin	associated with	llus caldotenax DNA p	earothermophilus DNA poly	5 standard deviations	•	ar beet chitinase 1.	coded by clone	Monkey erythropoietin.
524	300	303	266	241	205	167	167	167	166	166	134	134		above mean	2037	1594	1267	859	516	428	392	311	204	200	188	188	156	106	1048	635	405	360	355	119	119	900	877	76	above mean	350	439	193	192
15	л C	ı U	15	15	15	15	15	15	15	15	15	15	15	***	16	16	16	16	16	16	16	16	16	16	16	16	16	16	17	17	17	17	17	17	17	18	18	18	***	19	20	20	20
48	1 C	63	42	31	26	38	38	38	38	38	28	28	27		61	59	66	72	53	44	45	48	43	41	39	39	36	25	64	33	71	39	39	30	30	64	60	60		60	40	44	43
4.61		4.61	٠	٠		٠		4.61	٠			4.61			5.03														.4	.4	.4	٠.	٠.	٠.	5.45	•				6.29			
00	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0

US-08-223-263-1 (1-353) R29164 PRP3.

R29164 standard; Protein; 121 AA. R29164; R29164; 21-APR-1993 (first entry) PRP3

FT FT FT Proline-rich protein; PRP1; PRP2; PRP3; PRP4; PRP3g12; pistil; flower; sterile; plant; hydroxyproline; glycoprotein; gum. Nicotiana alata.

Key Location/Qualifiers
Region 23..24
/note= "Ser-Pro repeat unit"

*

ì

Salder.

Listing for Mary Hale

Tue Apr 11 13:27:35 1995

4

SSS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	1 T T
"Ser-Pro 13 "Ser-(Pro) 3 "Ser-(Pro) 5 "Ser-(Pro) 6 "Ser-Pro 16 "Ser-Pro 17 "Se	Region 2728 /note= "Ser-Pro repeat unit" Perion 31 32

Tue Apr 11,13:27:35 1995

Page Ġ

Initial Residue Gaps PSPSPPPSPSPSPSPSPPADDMAPSPSP-AAAPTPPDMIIRPSARLAGNMVISCIH
60 70 80 90 100 110 120 230 X 240 250 260 RAKIPGLLNQTSRSLDQIPGYLNRIHELLNGTRGLFP-300 310 320 330 340 350 XSPSPTHPTGQYTLFPLPTPTVQLHPLLPDPSAFTPTPTSPLLNTSYTHSQNLSQEG Identity Score GTESL--PDATIOR-FKFKWPFFGKSPKNSPKSSPSRSPPPKREQPSPPPPVKSPPPPSPPP X 10 20 30 40 50 0 11 22 28% 11 Matches = 35 Conservative Substitutions Optimized Score () II -GPSRRTLGAPDISSGTSDTGSLPPNLQPG 35 270 Mismatches Significance = 280 .54 78 0 290

US-08-223-263-1 (1-353) R29165 PRP3 (from genomic clone PRP3g12). R29165 standard; Protein; R29165;

151 Ā

gum. pistil;

Region /note= PAPR-1993 (first entry)
21-APR-1993 (first entry)
PRP3 (from genomic clone PRP3g12).
Proline-rich protein; PRP1; PRP2; PRP3; PRP4; PRP3g12;
Proline-rich protein; PRP1; PRP2; PRP3; PRP4; PRP3g12;
Proline-rich protein; Nydroxyproline; glycoprotein; g Region Region Region /note= /note= note= 'label= sig_peptide "Ser-(Pro)4 repeat unit" 81..85 "Ser-Pro "Ser-Pro "Ser-Pro "Ser-(Pro)3 repeat unit" repeat unit* 61..62 repeat unit*
57..58 repeat unit* 65..68 1..23

Region /note= Region /note= Region Region Region Region note= "Ser-(Pro)4 repeat unit" 98..99 "Ser-(Pro)4 repeat unit" 105..106 "Ser-(Pro)4 repeat unit" 93..97 "Ser-(Pro)4 repeat unit" "Ser-Pro "Ser-Pro repeat unit"
100..104 repeat unit"
107..111 112..114

Region /note=

*Ser-Pro

repeat unit* 88..92

Region

note=

"Ser-(Pro)

b)4 repeat unit 86..87

Region

note=

"Ser-Pro repeat u 123..124 *Ser-(Pro)2 repeat unit* 121..122 unit"

10

Listing for Mary Hale

Tue Apr 11 13:27:35 1995

6

```
Initial Score
Residue Identity
Gaps
                                                                  15-MAY-1992; AU0222.
15-MAY-1991; AU-006159.
(UYME) UNIV MELBOURNE.
Chen C, Clarke AE;
WPI; 92-415710/50.
                                                                                                                                                                                                                                   Claim 2; Pag 47 + Fig 5; 57pp; English.

The sequence below is derived from the genomic clone PRP3g12.

The polypeptide of 15,645 D is rich in proline (35%) and serine (16%). The sequence comprises 6 Ser-(Pro)4, 1 Ser-(Pro)3, 1 Ser-(Pro)2 and 8 Ser-Pro repeats.

1 Ser-(Pro)2 and 8 Ser-Pro repeats.

The sequences of the PRP1, PRP2, PRP3 (cDNA and genomic DNA), PRP4 genes are given in Q31794-98. The for nietilenestic series are given in Q31794-98. The sequences may be use
                                                                                                                                                                                                                                                                                                                                                                      N-PSDB; Q31797.

Isolated DNA sequence encoding proline rich proteins of Nicotiana elaman - used to express useful proteins and identify regulatory elements for pistil specific expression, to give e.g. longer
                                                                                                           glycoproteins. These could by used as effective substitutes for gums, e.g. guar gum and gum arabic, in foods and for non-food applications, where emulsification, thickening and stabilisation are required.
                                                                                                                                                           PRP4 genes are given in Q31794-98. The DNA sequences may be used to identify regulatory elements responsible for pistil-specific expression. Such elements may be used to provide pistil-specific expression of foreign or heterologous genes, e.g. to enable longer flower life or in the creation of male- or female-sterile plants. The PRPs may be converted by hydroxylation and glycosylation into hydroxyproline-rich
                                                             Sequence 151 AA;
10 A; 5 R; 4 N; 1
3 I; 12 L; 8 K;
                                                                                                                                                                                                                                                                                                                                                           flower life
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       26-NOV-1992.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         NO9220713-A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       note= "Ser-Pro repeat unit"
   20
24%
11
                                                                 м;
50
Optimized Score = 34
Matches = 38
Conservative Substitutions
                                                                 ΉB
                                                             .
45
                                                                 ئة <del>(</del>3
                                                               3 Q;
26 S;
                                                                  ωa
                                                                  ΗE
                Significance
Mismatches
                                                                  10
                                                                  Σ.
                                                                  0 5
                                                                 ۲.
ن
                                                                  σN
                                   11
      11 (1
                                                                 <.#
                                   9
   104
104
```

RTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRIHELLNGTRGLFP-----G

3. US-08-223-263-1 (1-353) P50301 Monkey erythropoietin.

PDMLLRPSARLAGNMVISGLE 140 150

AC ID P50301 standard; protein; 192 AA P50301;

, a Same of the same o

Listing for Mary Ha Ø

Tue Apr 1 13:27:35 1995

Page

P50343 standard; Protein; 193 A

0-MAR-1992

Gaps Residue Initial SKEDACI US-08-223-263-1 (1-353) P50343 EPO encoded by clone lambda-HEPOFL13. VNFYAWKRMEVGQQAVEVWQGLALLSEAV--LRGQAVLANSSQPFEPLQLHMDKAISGLRSITTLLRALGAQ X 10 20 30 40
MELTELLIVVMLLL-TARLTLSSP---APP--ACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTPVLLPAVD EAISLPDAASAAPLRTITADTFCKLFRVYSNFLRGK---LKLYTG-EACRRGDR 150 160 170 180 190 X 130 140 150 160 170 180 -LOSILGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPPTTAVPSRTS FSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQL--MGVHECPAWLWILLSIVSIPIGLFVPGAFFRLICDSRVIERVILEAKEAENVTMGCSESCSINENITVFDTK X 10 20 30 40 50 60 70 11-DEC-1984; U02021.
13-DEC-1983; US-561024.
21-FEB-1984; US-6582185.
28-SEP-1984; US-655841.
30-NOV-1984; US-65298.
(KIRI-) KIRIN-AMGEN INC.
WPI; 85-159229/26. Sequence 192 20 A; 13 R; 6 6 I; 30 L; 7 New polypeptide having properties of erythropoletin - is prepd. by cultivation of transformed eucaryotic or procaryotic host Disclosure; Page 38; 113pp; English.

Monkey erythropoletin encoded by this sequence is essential for red blood cell formation and is used for the diagnosis and treatment of blood disorders such as anaemia. Large amounts of EPO may be obtained using recombinant DNA techniques in contrast to small amounts obtained from plassma and urine. See also N50345-47, N50349-50 and P50298-P50300. N-PSDB; N50348 Monkey. W08502610-A. disorder; Erythropoietin; red blood cell; erythrocyte; anaemia; blood; 20-JUN-1985. Monkey erythropoietin. 01-JAN-1980 Erythropoietin. Homo sapiens. EPO encoded by clone lambda-HEPOFL13 250343; Identity Score ds. 80 11 11 11 (first entry) (first entry) X X 4 6 20 27**%** 27 90 Optimized Matches K D ** Conservative Substitutions 50 'T D 100 11 Score ΑĊ. 7 14 H H ŝö 110 3 9 13 54 54 HE 40 Mismatches Significance ΣZ 12 4 SG---QVRLL--LGA-Υ. Θ. 13 120 $\Pi = \Pi$ II. < ;: 6.71 117 0

> Listing for Mary Hale

i.

Tue Apr 11 13:27:35 1995

Page

Ø

Initial Residue Gaps VNFYAWKRIEVAQQAVEVWQGIALLSEAV--LRGQALLVNSSQPWEPLQLHVDKAVSGLRS-LTTLLRPLGA 80 90 100 110 120 130 140 QKEAISPPDAASAAPLRTITADTFRKLFRLYSNF--LRGKLKL 150 160 170 180 140 RTTAHKDPNA---IGVHECPASIMILISIPIGLPVRGAPPRLICDSRVLLERYLLEAKQAENITTGCAEHCSINENIIVPDTK X 10 20 30 40 50 60 70 ${ t FSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQG-$ Sequence 20 A; 14 8 I; 33 Vector system for introducing heterologous DNA into eukaryotic cells - comprises prod. gene and accessory DNA for enhanced expression of heterologous protein by the cells. Disclosure; Fig 13; 62pp; Danish.

See also US4740461 88.04.26 (8819) (first major country equivalent). The sequence is encoded by clone lambda-HEPOFI.13 which contains the gene encoding erythropoietin obtd. from a human foetal DNA library. The gene can be expressed using a novel vector system dislosed in the specification.

See also P50342.

Sequence 193 AA; N-PSDB; N50443 DK8406107-A WPI; 85-318061/51. (GENE-) GENETICS INST. Kaufman RJ. 22-JAN-1985; 22-JAN-1985; 21-DEC-1984; 04-DEC-1984; 27-DEC-1983; 27-DEC-1983; 27-DEC-1984; Peptide /label= .6-AUG-1985 'label= mature_EPO X 10 20
MELTELLLVVMLLLTARLTI-SSPAPP--ACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTPVLLPAVD Identity signal_sequence 27..193 1 R; 6 0 0 0 281862. US-565627. US-566057. ZA-010034. US-693258. US-688622. US-677813. ZZ. 20 26**%** 25 Location/Qualifiers 1..27 0 6 Optimized Matches K O Conservative Substitutions 40 'ii (ii 4 12 Score À C: 8 14 H 11 ŝ 10 44 53 ΗF ---YTGEASRTGDR 190 X Significance Mismatches 40 ₹,7 411 `. છે 3 12 11 11 11 < # 6 5.71 124 0

US-08-223-263-1 (1-353) Sugar beet chitinase 1.

H R28150 standard; Protein; 439 AA



'n

Sugar beet chitinase 1.
SBC-1; fungicide; anti-fungal agent;

extensine.

17-MAR-1993

(first entry)

Beta vulgaris cv monova.

Listing for Mary Hal

Ð

Tue Apr 11

13:27:35

1995

6

**

```
Residue
                                                                                                                                        Initial
                                                                                                                                                            180 X 190 200 210 CVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASAR--
DQIPGYLNRIHELLNGTRGLFPGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPT
                                                                                                                                                        Sequence
15 A; 33
19 I; 26
                                                                                                                                                                                               Sugar beet chitinase 4 and corresponding DNA - inhibits germination and growth of chitin-containing fungi and is used to produce genetically transformed plants claim 5; Page 164-168; 254pp; English.

A clone encoding the chitinase 1 gene was isolated from a sugar beet EMBL3 genomic library. The sequence encodes a protein having 439 amino acid residues. Transgenic plants having increased resistance to mematodes and chitin-containing plant pathogens, partic. fungi, can be produced using genetic constructs containing plant pathogens.
                                                                                                                                                                                                                                                                                                                                                                     the vacuole" WO9217591-A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                   chitinase 1 protein to the cell wall after modification of the prolines to
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           ХeУ
                                                                                                                                                                                                                                                                                                 N-PSDB;
                                                                                                                                                                                                                                                                                                         Berglund L, Bojsen K, Mikkelsen WPI; 92-366261/44.
                                                                                                                                                                                                                                                                                                                                       07-APR-1992; DK0108.
08-APR-1991; DK-000616.
                                                                                                                                                                                                                                                                                                                                                                                                 Region
                                                                                                                                                                                                                                                                                                                                                                                                                               extensines"
                                                                                                                                                                                                                                                                                                                                                                                                                                         glycosylated hydroxyprolines,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            Domain
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 Peptide
                                                                                                                                                                                       compositions.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                     /label= proline_rich
/note= "possibly involved
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         /label= hevein_domain
Domain 47..178
                                                                                                                                                                                                                                                                                                                             (DANI-) DANISCO AS.
                                                                                                                                                                                                                                                                                                                                                                                                 /label= functional domain Region 417..439
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     /label= leader
                                                                                                                                                                                                                                                                                                                                                                                        note= "probably"
                                                                                                                            Identity
                                                                                                                                      Score
                                                                                                                                                          33 R;
26 L;
                                                                                                                                                                                                                                                                                                 Q29965
                                                          MKIKTS-PSFLLGLICLALVLLLGEGVQCGRQCNTTDTNCLSGCSVGRPSRP X 10 20 30 40 50
                                                                                                                                                          439 AA;
16 N;
; 8 K;
                                                                                                                     0 0 0
                                       260
                                                                                                                                                                                                                                                                                                                                                                                                                       179..416
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            27..46
                                                                                                                                                                                                                                                                                                                                                                                        directs
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         Location/Qualifiers
                                                                                                                                                          <sup>3</sup> 13
                                                                                                                                                           K O
                                                                                                                    Conservative Substitutions
                                                                                                                                       Optimized
                                                                                                                                                            19
                                                                                                                                                                                                                                                                                                                                                                                         the
                                        270
                                                                                                                                                                                                                                                                                                                                                                                                                                                                      in anchoring
                                                                                                                                                           äά
                                                                                                                                                                                                                                                                                                                                                                                       protein
                                                                                                                                                          12 C;
109P;
                                                                                                                                                                                                                                                                                                                                                                                                                                          as in
                                                                                                                                        Score
                                                                                                                                                                                                                                                                                                                     'n,
                                       280
                                                                                                                                                                                                                                                                                                                      Nielsen
                                                                                                                                                          14
23
                                                                                                                                                                                                                                                                                                                                                                                         ç
                                                                                                                                a b
  90
                                                                                                                                                           SO
  220 230 240
-TTGSGLLKWQQGFRAKIPGLLNQTSRSL
                                                                                                                                                            21
39
                                                                                                                                                                                                                                                                                                                      <u>주</u>
                                                                                                                               40
45
                                                                                                                                                           ΗF
                                       290
                                                                                                                                                           90
                                                                                                                              Significance
Mismatches
                                                                                                                                                           ₹2
                                                                                                                                                           31
12
                                       300
                                                                                                                                                           સ્ છ
                                                                                                                                                           7
10
                                                                                                                                       II
                                                                                                                                                           ,< #:
                                                                                                                                        6.71
                                        310
```

Initial Residue

Score Identity

0 0 0

19 20% 66

Optimized Matches

Score

60 79

M:Si

ignificance smatches

H H

6.29 233 0

6

Conservative Substitutions

Sequence 44 A; 19 9 I; 53

T, 73

Z Z

13 D; 9 M;

22

7. B 13

79.0

9 18 ŝ

11

ΗE

13

Σ,2

16 21

₹6 12 21

< #

350 12

See also

MELTELLLVVMLLLTARLTLSSPAP-PACDL--RVLSKL---

---LRDSHVLHSRLSQCPEVHPLPTPVLLPA

50

8

40

VDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPP-

Claim 18; Fig 11A; SBp; English.

The protein sequence was deduced by screening adult mouse brain cDNA libraries with a 1.4kb GDF-1 sequence isolated from a mouse embryo CDNA library (see Q20688). The DNA sequence of 2.7kb obtd. comprised two ORFs, one of which was highly homologous to the 1.4kb probe.

(Q20688). Upstream of the GDF-1 coding region is a 2nd ORF, encoding mUGG-1 (Upstream Of GDF-1) which terminates upstream of the GDF-1 ATG codon. Due to the presence of multiple stop codons in the region between mUGG-1 and mGDF-1, at least 4 frameshifts would be required to translate the two ORFs as a single protein. Hydropathicity analysis of the amino acid sequence of mUGG-1 revealed multiple clusters of hydrophobic residues, indicative of membrane spanning domains. mUGG-1 does not contain an obvious N-terminal signal US-08-223-263-1 (1-353) R20229 UOG-1. 320 330 340 350 x LPTPVVQLHPLLPDPSAPTPTPTSPLLNTSYTHS--QNLSQEG WPI; GDF-1 DNA segment encoding a mammalian GI tumours, birth defects and genetic N-PSDB; Q20689 14-JUN-1991; U04096. 15-JUN-1990; US-538372. 16-NOV-1990; US-614452. 09-JAN-1992. WO9200382-A. 06-MAY-1992 R20229 standard; Protein; (CARN-) CARNEGIE INST sequence 92-041563/05. ; mouse; R20228, 30 and R22375. (first murine; hydrophobic clusters entry) WASH 350 Ą GDF-1 protein - ic diseases. 160 for diagnosis of



Ø

494

. (

*

*

Х,

3

Tue Apr 11 13:27:35 1995

Page 11

US-08-223-263-1 (1-353)

B.stearothermophilus DNA polymerase

R45155 standard; Protein; 876 AA R45155; 16-JUN-1994

16-UUN-1994 (first entry)
B.stearothermophilus DNA polymerase.
Bacillus stearothermophilus; DNA polymerase; thermal stability;
heat resistant; plasmid pUIF101.

Bacillus stearothermophilus (IAM11001).

J05304964-A.

9-NOV-1993

WPI; 93-408323/51. N-PSDB; Q54170. 27-APR-1992; 131400. 27-APR-1992; JP-131400. (TAKI) TAKARA SHUZO CO LTD

DNA polymerase gene - is isolated from plasmid pUIF101, useful as reagent for gene engineering research Example 2; Page 8-11; 11pp; Japanese.

DNA was extracted from Bacillus stearothermophilus IAM11001 and subjected to PCR amplification using the primers Q54168 and Q54169. A heat-resistant polymerase (R45155) can be recombinantly produced by transforming E.coli host cells with the gene comprising sequence by transforming E.coli host cells Q54170.

Sequence 83 A; 55 48 I; 99 876 AA; 25 N; 62 K; 49 22 K O 330 ኳ 🖰 37 P,C 38 37 ŝ 90 40 HE 50 ₩, 2, 44 28 ∵. 17 62 < ;:

Initial Score II Optimized Score Matches 11 11 60 84 Mismatches Significance H 5.87

Residue Identity 11 11 18 21**%** 69 Conservative Substitutions

.....

Listing for Mary Hale

â

Tue Apr 11 13:27:36 1995

Page 12

RPFIDELRNEQDRLLVELEQPLSSILAEMEFAGVKVDTKRLEQMGEELAEQLRTV 460 470 X 510 320 TP-VVQLHPLLPD-PSAPTPTPTSPLL---340 350 x

410

430

US-08-223-263-1 (1-353) R28348 Bacillus caldotenax DNA polymerase

AA.

R28348 standard; Protein; 877 R28348;

07-APR-1993 (first entry)
Bacilius caldotenax DNA polymerase I.
PolI; polymerase chain reaction; stra
plasmid pUI101.

strain YT-G; DSM406;

Bacillus caldotenax. location/Qualifiers

EP-517418-A. /note= "corresponds

to TTG codon"

Misc_difference 1

27-MAY-1992; 304763. 03-JUN-1991; JP-157368. 07-NOV-1991; JP-318685. 24-FEB-1992; JP-072090. 25-FEB-1992; JP-073161. 06-APR-1992; JP-112400. (TAKI) TAKARA SHUZO CO L LTD.

* : . . . 240

Fujita K,

ta K, Ishino Y, Kato I, Uemori 92-408872/50.

Η

```
Residue Identity
                                                                                                                                                                                                                                                                                                                                                         Initial
                                                                                                                                                                                                                                                                                                                                                                             IGEKTAVKILROFGTVENVLASIHEIKGEKIKETLROHREMA---ILSKKIAAIRRDAPVEISIDDIAYOGE 200 210 220 230 240 250 260
                                                                                VVNEHGRFFLRPETALADPQFVAWLGDETKKKSMFDSKR---AAVALKW---
340 350 360 370 380
                                                                                                               V---PSRTSL-VLTLNELPNRTSGL-LET---NFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLDQIPG
                                                                                                                                              DREKVVALFKELGFQSF LEKMESPSSEEEKPLAKMAFTLADRVTEEMLADKAALVVEVVEENYHDAPIVGIA 270 280 290 300 310 320 330
                                                                                                                                                                   120 130 140 150 160 170
VRLLIGALQSILGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKVRFIMIVGGSTLCV----
                                                                                                                                                                                                                                          HPLPTPVLL--
                                                                                                                                                                                                                                                                       GFEVKVISGDRDLTQLASPHVTVDITKKGITDIEPYTP-EAVREKYGLTPEQIVDLKGLMGDKSDNIPGVPG
130 X 140 150 160 170 180 190
                                                 YLNRIHELLNGTRGLFPGPSRRTLGA-
                                                                                                                                                                                                                                                                                                                                                                          Sequence
84 A; 54
48 I; 99
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Gene
                                                                                                                                                                                                                                                                                                                                                                                                See also R28340-R28347 and Q31651.
Sequence 877 AA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         N-PSDB; Q31650.

Gene for Poll type DNA polymerase and cloning method -
                                                                                                                                                                                                                                                                                                                                                                                                                     polymerase deduced to have the amino acid sequence given here.
                                                                                                                                                                                                                                                                                                                                                         Score
                   --LLLAAYLLDPAQGVDDVAAAAKMKQYEAVRPD-
400 410 420
                                                                                                                                                                                                                                                                                                                                                                           54 R; 25
99 L; 62
                                                                                                                                                                                                                                                                                                  X 10 20 30 40 MELTEL-LLVVMLLLT-ARLTLSSPAPPACDLRVLSKLLRDSHVLHSRL-
                                                                                                                                                                                                                                                                                                                                      II A
                                                                                                                                                                                                                                                                                                                                                        II
                                                                                                                                                                                                                                         --PAVDFSLGEWK-TOMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQ
                                                                                                                                                                                                                                                                                                                                                                           7 AA;
5 N; 48 D;
2 K; 22 M;
                                                           270
                                                                                                                                                                                                                                                                                                                                   Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                              Matches
                                                                                                                                                                                                                                                                                                                                                        Optimized Score
                                                                                                                                                                                                                                                                                                                                                                             330
                                                                                                                                                                                                                                                                                                                                                                             ω
₩,₩
                                                                                                                                                                                                                                                                                                                                                                           2 C;
39 P;
                                                                                                                                                                                                                                                     80
                                                                                                                          210
                                               PDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTL
                                                                                                                                                                                                                                                                                                                                                                             39
36
                                                                                                                                                                                                                                                                                                                                                         11
                                                                                                                                                                                                                                                                                                                                                                             φÓ
                   -EAVYGKGAKRAVPDEPVLAEHLVRKAAAI
430 440 450
                                                                                                                                                                                                                                                     90
                                                                                                                        220
                                                                                                                                                                                                                                                                                                                                                                              90
39
                                                                                                                                                                                                                                                                                                                                               96
03
350
                                                                                                                                                                                                                                                                                                                                                                             크
                                                                                                                                                                                                                                                                                                                                               Significance
Mismatches
                                                                                                                                                                                                                                                                                                                                                                             ,
O O,
                                                                                         ---KGIEL-
                                                                                                                                                                                                                                                    100
                                                              290
                                                                                                                                                                                                                                                                                                                                                                             ₹ 2
                                                                                                                                                                                                                                                                                                                                                                             44
28
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                for the
                                                                                                                                                                                                                                                                                                                                                                             Υ.Θ
                                                                                          -CGVSFD----
                                                                                                                                                                                                                                                                                                                                                                             18
62
                                                                                390
                                                                                                                                                                             -RRAP-PTTA
                                                                                                                                                                                                                                                                                                     --SQCPEV--
                                                                                                                                                                                                                                                                                                                                                          n
                                                              300
                                                                                                                          240
                                                                                                                                                                                                                                                                                                                                                                             < μ
                                                                                                                                                                                                                                                                                                                                                          <u>ر</u>
                                                                                                                                                                                                                                                                                                                                              5.87
235
```



FPL-PPTL--PTPVVQLHPL--LPDPSAPTPTPTSPLLNTSYTHSQNLSQEG

GDRERERGMEIS--

-PMCDKHTASVEKSQVGFIDYIVHPLWETWADLVHPYSAIRQSPSPPPEEESRGF

" Listing for Mary Hale

Tue Apr 11 13:27:36 1995

74

Initial Residue Complementary screening for genes and prods. - e.g. RAS protein and cAMP, that modify, complement or suppress genetic defect and correct associated phenotypic alteration

Torrect associated phenotypic alteration

Siclosure; Page 145; 169pp; English.

In the specification this sequence is given the SEQ ID NO. 2 and is additional to SEQ ID NO. 3: 1-55, i.e. the specification contains two sequences denoted as SEQ ID NO.2: The only reference to SEQ ID NO.2 in the text is to a 10 base pair linker. The origin and identity of R14854 is therefore obscure. Other sequences

In the specification were isolated from human glioblastoma cells and encode cyclic nucleotide PDEs and RAS-related pypeptides. (They were isolated by their ability to complement or suppress genetic defects in a biochemical pathway involving cAMP or which is controlled by a RAS protein).

Sequence 900 AA;

80 A; 63 R; 27 N; 36 D; 0 B; 12 C; 47 Q; 68 E; 0 Z; 37 G; 27 H; 89 P; 85 S; 45 T; 13 W; 16 Y; 49 V; US-08-223-263-1 (1-353) R14854 Protein associated with biochemical 120 130 140 150 160 170 180 GQVRLLLGALQSLLGTQLPPQGRTTA--HKDPNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPP--TTAVP ADLKTMVETKKV-TSSGVLLLDNYSDRIQ-LINTNSELALMYNDESVLENHHLAVGFKLL-470 480 490 HPLPTPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQ-Wigler MH, Colicelli JJ; WPI; 91-339841/46. N-PSDB; Q14645. 20-APR-1990; 31-OCT-1991. Homo sapiens. WO9116457-A. Protein associated with biochemical pathway involving cAMP. RAS; oncogene; cancer; cyclic nucleotide phosphodiesterase. R14854 standard; Protein; 900 30-JAN-1992 (first entry) R14854; (COLD-) COLD SPRING HARBOR. 19-APR-1991; U02714 Identity MELTELLLVVMLLLTARLTLSSPAPPACD-LRVLSKLLRDS-0 11 0 US-511715 18 20% 74 Optimized Matches Conservative Substitutions 20 --QEDNCDIFQNLSKRQRQSLRKWVIDMVLATDMSKHMTLL 500 510 520 ₽. Score -VLRNMVHCA-DLSNPTKPLELYRQWTDRIMAEFFQQ 560 100 64 82 pathway involv 40 Significance Mismatches 110 -HVLHSRLS-QCPEV 0 0 S 5.87 239 0 590

× ... Tue Apr 11 13:27:36 1995

Page 15

10. US-08-223-263-1 (1-353) R12538 SMUC-87 in

SMUC-87 intestinal mucin CDNA clone

02-SEP-1991 R12538 standard; Protein; 119 AA. SMUC; SMUC-87 intestinal mucin Region Homo sapiens. Region 'label= repeat 'label= repeat intestinal mucin; diagnosis; cancer. Afirst entry) 74..97 51..73 28..50 Location/Qualifiers cDNA clone

/label= repeat W09108217-A. Region 'label= repeat 98..119

13-JUN-1991.
04-DEC-1990; U07087.
05-DEC-1989; US-447140.
(REGC) UNIV OF CALIFORNIA.
Kim YS, Gum JR;
WPI; 91-193147/26. see also Q12217-21 and Q12247-51. Sequence 119 AA; cione smuc v/ contains 4 tandem repeats. It is notable in possessing homology segments of 24 amino acids and of 22 amino acids, demonstrating that these conserved segments may be of variable length, having amino acid insertions and deletions. The extra amino acid in the third repeat is a methionine (residue 85) while a threonine is absent from the fearth second. Disclosure; Fig 2(E); 57pp; English. Clone SMUC 87 contains 4 tandem repeats. 1 homology segments of 24 amino acids and of N-PSDB; Q12221. Nucleic acid encoding human intestinal for producing prods. for diagnosis and such as cancer À Ç G mucin peptide(s) - used
treatment of disorders Ö -H 0 2 N ç; _ H

Listing for Mary Hale

Tue Apr 11 13:27:36 1995

Page 16

11. Residue Initial SO US-08-223-263-1 (1-353) R07674 Intestinal 띥 TRITTINIVIPTRIPTRIGTOTRIPTRITTITIMVIRIPRITS (QTRIPTRIETZ) 60 70 80 90 100 LKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRIHEL-œ Human intestinal mucin DNA, diagnosis and treatment of ε colitis. Disclosure; Fig 4E; 45pp; English.

The sequence was deduced from a clone isolated from a cDNA library using clone SMUC 41 as a probe. The DNA sequence allows determination of the primary amino acid structure, difficult to characterise biochemically because the protein backbone is so heavily glycosylated and conditions required to remove the carbohydrate tend to result in breakage of the protein. Prods. of the DNA can (REGC) University of California. Kim YS; Gum JR WPI; 90-348495/46. 25-JUL-1989; U03206. 14-APR-1989; US-338710 01-NOV-1990 Region 98.. /label= 22 aa repeat Region Intestinal mucin deduced from clone SMUC 87. Epithelial cancer; cystic fibrosis; ulcerative glycosyl transferase; familial polyposis coli; R07674 standard; protein; 119 R07674; 290 N-PSDB; Q06487 WO9012892-A. Region Region Homo sapiens. /label= /label= /label= 22-FEB-1991 (first entry) lycosylation patterns. Identity Score Į; N 230 24 KFGL. 23 23 ۲. 300 aa repeat 98..119 aa 0 0 Intestinal mucin deduced w repeat 74..97 repeat 51..73 ζ. 17 27% 21 28..50 Location/Qualifiers Optimized Matches ĭ Conservative QIIQKFKPIFFLSFTATTTTTENPTPTPTTTTTTTTTTPTPTSTQSTTP G Ŧ e.g. cancers, cystic fibrosis and B A 23 320 Score Ρ, Substitutions S from clone SMUC --LNGTRGLFPGPSRRTLGAPDISSGTSDTGSLP s; 56 T; Sea 30 35 colitis; Significance Mismatches 0 TTVTPTPTPTS-THT ξ. 270 -TSPLLNTSYTHSQNLSQ 0 40 340 ť; 4 0 0 0 < 'n X 350 S .45 73 0



×.

s. (g

8

. X

Initial Residue 999999888 290 300 310 320 330 PNLQPGYSPSPTHPPTGQYTLFPLPPTLPTPVVQLHPLLPDPSAPTPTP-TPITTTNTVTPTPTPTGTQTPTPTPITTTTTMVTPTPTITSTQTPTPTPTITTTTVTPTPTPTS-IHT 60 70 80 90 100 110 X 20 230 240 250 260 270 280 LINGGERAKIPGLINGTESLPULAPITAGESLP Sequence 1 A; 0 8 I; 2 be used to produce ADS userum the transportation patterns or glycosyltransferase activity. They can be used for the early detection and differential diagnosis of cancers esp. epithelial cancer, and also in diseases with altered intestinal mucin prodn. such as cystic fibrosis and colitis. See also R07670-76. Identity Score e 119 AA; R; 2 N; 0 L; 3 K; 1 KFGL---0 0 0 17 27**%** 21 K O Conservative Matches Optimized **5**0 ΉØ 23 Score P.C Substitutions II II ωÖ 1 56 30 35 H.E Significance Mismatches 00 Σ. TSPLLNTSYTHSQNLSQ 0 1 ۲; و; بر ي 11 II 11 < ;: G x 350

EG

12. US-08-223-263-1 (1-353) R33420 Human IL-8

IL-8 receptor from clone Ŗ

standard; Protein; 355 A

26-JUL-1993 (first entry)

Human II-8 receptor from clone p2.
Interleukin-8 receptor; probes; ge
intracellula-calcium mobilising; l gene therapy; gro receptor;
; ligand-binding; MIP-2 rece

MIP-2 receptor.

WO9306229-A. Homo sapiens.

01-APR-1993.

14-SEP-1992; U07641. 13-SEP-1991; US-759568. (USSH) US DEPT HEALTH HUMAN SERVICE

(USSH) US DEPT HE Murphy PM; WPI; 93-117549/14.

N-PSDB; Q38747

New interleukin-8 receptor aminoacid sequence - and corresp. cDNA expressed in Xenopus laevis oocytes or transfected host cells, for screening ligands of IL-8 receptor and gene therapy Claim 1; Fig 3; 39pp; English.

cDNA libraries from 2 and 3.5 kb fractions of poly(A)+ RNA from HL60 neutrophils sepd. of a sucrose gradient were made in UnizAP. The libraries were screened with F3R oligonucleotide probe (from rabbit II-8 receptor) and under low stringency with a p2 cDNA probe synthesised from random primers, to isolated the clone p2, encoding human II-8R or related MIP-2 receptor may be detected using portions of the p2 clone as probes. P2 may also be used to screen

No. -3.1.4 * 40 235 á Ť

> Listing for Mary Hale 13:27:36

Tue .

1995

Page

18

Apr 11

333888 333888 for ligands of II-8R and may also patient deficient in II-8R. The intracellular calcium-mobilising Sequence 355 AA; 22 A; 18 R; 13 N; 13 D; 0 B; 9 25 I; 65 L; 13 K; 8 M; 21 F; 13 also be used IL-8R is a gro rece and ligand-binding in gene therapy to a gro receptor and properties. treat a has

ΑĊ 30 S O 11 18 ΗE 60 Σ 2 17 14 **∺** છ 25 < H

Initial Residue Identity Score 11 II 17 23% 51 Conservative Substitutions Optimized Score 11 (1 39 55 Significance Mismatches 11 11 11 G 5.45 125 0

QETCERRNHIDRALDATEILGILHSCINP-280 290 300 QLSGQVR---120 -LLLGALQSL--LGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKVRFIMLVGGSTLCVRRAPP 40 -LIYAFIGQKFRH--310 150 -GLLKILAIHG---LISKDSLP

180 190 200 X 210 TTAVPSRTSLVLTINELPNRTSGLLETNFTASART -KDSRPSFV----GSSSGHTSTTL 340 350 X

13.

US-08-223-263-1 (1-353) R28273 Sequence in D low affinity human inter

standard; Protein; 360 B

04-APR-1993

(first entry)

Sequence in a low affinity recombinant human interleukin-8

receptor polypeptide; G-protein-coupled

(IL-8) receptor polypeptide in 4AB

Homo sapiens. WO9218641-A.

10-APR-1992; U02977. 10-APR-1991; US-685101. 09-JUL-1991; US-726606. 09-DEC-1991; US-803842. (REPK) REPLICEN CORP.

(REPK) (UYBO-) UNIV BOSTON

Navarro J, Thomas KM, Witt WPI; 92-382123/46.

뮭,

N-PSDB; Q30012.

Recombinant mammalian interleukin-8 receptor - interleukin-8 binding antagonists, used to treadisclosure; Fig 2; 71pp; English. used for screening



Tue Apr 11 13:27:36

1995

14. Residue Identity Initial 88888888888888888888888888 SXEDACU 180 190 ZUU A ZIV TTAVPSRTSLVLTLNELPNRTSGLLETNFTASART US-08-223-263-1 (1-353)
R14838 Protein deduced from human QETCERRNHIDRALDATEILGILHSCLNF-290 300 310 QLSGQVR----LLLGALQSL--LGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKVRFIMLVGGSTLCVRRAPP RILPQ-SFGFIVPLLIMLFCYGFTLRTLFKAHMGQKHRAMRVIFAVVLIFLLCWLPYNLVLLADTLMRTQVI ${\tt SRLSQCPEVHPLPTPVLLPAVDFSL-GEWKTQM-EETKAQDILGAVTL--LLEGVMAARGQLGPTCLSSLLG}$ DRYLAIVHATRTITOKRYLVKFICLSIWGLSLLLALPVLLFRRTVYSSNVSPÄCYEDMGNNTANWR--150 X 160 170 180 190 200 ---KDSRPSFV---GSSSGHTSTTL 350 360 kb in size. This insert was sequenced. The protein deduced from the FIR clone demonstrates that it belongs to the family of G-protein-coupled receptors. The deduced protein sequence indicates seven putatuve transmembrane segments. A human peripheral blood leukocyte lambda gtl1 cDNA library (5' stretch) was screened with a 652 bp EcoRI/BamHI fragment (including nucleotides 27 to 625) of the rabbit FIR clone. After tertiary screening several human clones which hybridized to the rabbit IL-B probe were isolated. The insert of one such clone, termed 4AB was sequenced (030012). Rabbit high affinity II-8 receptor gene was isolated from rabbit peritoneal neutrophils and used as a source of poly(A)+ RNA, to produce a rabbit neutrophil cDNA library. 250,000 recombinant plaques were screened for those which hybridized to an antisense oligonucleotide (Q30015). This probe was designed based on the sequence derived from the second transmembrane domain of G-protein-coupled receptors. After tertiary screening, six plaques were isolated. The insert was sequenced. The protein deduced from the second transmembrane domain deduced from the state of the sequenced. The protein deduced from the case of the sequenced of the sequenced of the protein deduced from the case of the sequenced. Sequence R14838 standard; Protein; 405 AA. R14838; Score 18 R; 14 65 L; 13 MELT--360 AA; R; 14 N; 14 D; L; 13 K; 9 M; 0 0 60 130 Conservative Substitutions Optimized Score 22 240 10 20 30 40 ---ELLLVVMLLLTARLTLSSPAPPAC--DLRVLSKLLRDSHVLH in in 140 9 13 --LIYAFIGQKFRH---GLLKILAIHG---LISKDSLP P,C 90 30 B II Glioblastoma cell cDNA. s o 150 12 18 90 39 55 ΗE 260 Significance Mismatches 60 ≅ 2 17 14 100 ₩. 25 II łł 5.45 125 was

30-JAN-1992 (first entry)
Protein deduced from human Glioblastoma cell cDNA.

oncogene; cancer; cAMP. sapiens.

310

320

Initial Score Residue Identity Complementary screening for genes and prods. - e.g. RAS protein and cAMP, that modify, complement or suppress genetic defect and cAMP, that modify, complement or suppress genetic defect and correct associated phenotypic alteration

Example 1B; Page 70; 169pp; English.

In the specification the sequence from which this amino acid sequence was deduced is given the SEQ ID NO. 13 and is described as the cDNA insert of plasmid pJC99. The insert sequence shows no significant homology to previously isolated genes. Plasmid pJC99 in E.coli (ATCC 68599) is described as the cDNA insert of the sequence shows no significant homology to previously isolated genes. Plasmid pJC99 in E.coli (ATCC 68599) is described as 39 (COLD-) COLD SPRING HARBOR. Wigler MH, Colicelli JJ; WPI; 91-339841/46. 31-OCT-1991. 19-APR-1991; U02714 20-APR-1990; US-511 the specification to particular sequences caution is advised in accepting that this containing a human glioblastoma cell cDNA insert encoding a RAS-related polypeptide. N.B. many of the references in the text of N-PSDB; Q14626. WO9116457-A. 36 R; 2 N; 9 74 L; 10 K; 6 MELT-11 11 11 D; 0 м; 7 Optimized Score = 71
Matches = 90
Conservative Substitutions ELLLVVMLLLTARLTLSSPAPPACDLRVLSKLLRDSHVLHSRLSQCPEV 7 B 33 ۲. ت 40 40 s 0 25 E; 0 18 T; 2 sequence is indeed pJC99 ! are incorrect, therefore Significance = Mismatches = Σ. 32 G; 8 Y; 10 H; 16 V; H 11 5.45 196 0

HPLPTPVL--

-LEGVMAARGQLGPTCLSSLLG

Tue Apr 11 13:27:36 1995

5 Page 21

LPPTLPTPVVQLH

RGL
X

15. US-08-223-263-1 (1-353)
R23970 MPLV env p

MPIV env protein with growth factor receptor prope

Initial Score = Residue Identity = Gaps = -R N-PSDF Q26674.

R N-PSDF Q26674.

Prolypeptides similar to v-mlp protein of MPIV - for diagnosis and Prolypeptides similar to v-mlp protein of MPIV - for diagnosis and T treatment of myeloproliferative diseases

T treatment of myeloproliferative diseases

S Claim 1; page 28; 75pp; French.

This polypeptide is encoded by clone MPIV107, isolated from a genomic bank prepared from clones of Mus dunni cells containing a unique copy of the MPIV provirus. The bank was screened with two 300bp long clones specific for MPIV env region. Restriction enzyme analysis showed that clone MPIV107 contained the entire MPIV genome except for the 3'LTR.

See R23971 and Q24674-7.

See R23971 and Q24674-7.

Sequence 635 AR;

Sequence 635 AR;

12 N; 23 D; 0 B; 22 C; 41 Q; 38 E; 0 Z; 36 G; 19 H; 17 I; 78 L; 12 K; 8 M; 18 F; 60 P; 64 S; 39 T; 21 W; 19 Y; 28 V; 50 60 70 80 90 100
LHSRLSQCPEVHP-----IP-----TPVILIPAVDFSLGEWKTQMEETKAQ-DILGAVTILLIEGVMAARGQLG VETATETAWISLUTALHIV--LGLSAVLGILLLRWQFPAHYRRLRHALWPSLPDLHRVLGQYLRDTAALSPP 490 X 500 510 520 530 540 550 PT--CLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKD KATVSDTCEEVEPSLLEILPKSSERTP--LP----LCSSQAQMDYRRLQPSCLGTMPLSVCPPMAESGSCC 560 570 580 590 600 610 (INRM) INSERM INST NAT SANTE & RECH MEDICALE. Charon M, Gisselbrecht S, Penciolelli JF, Souyri Tambourin P, Varlet P, Vigon I, Wendling F; WPI; 92-167154/20. 05-NOV-1992 (first entry)
MPIV env protein with growth factor receptor properties.
Myeloproliferative leukaemia virus; haematopoietic cell.
Myeloproliferative leukaemia virus. Region 449..635 /note= "gp70 of MPLV" W09207074-A. R23970 standard; protein; 635 AA. R23970; 30-APR-1992. .9-OCT-1990; WO-F07623. 9-OCT-1990; F00762 × MELTELLLVVMLLLTARLTL----SSPA----17 Optimized Score
28% Matches
57 Conservative Sub 10 ocation/Qualifiers Conservative Substitutions 130 11 - 11 33 48 -PPACDL-RVLSKLLRD----SHV Significance Mismatches ĸ U 5 40

Listing for Mary Hale

Tue Apr 11 13:27:36 1995

raye

22

```
Ouery sequence being compared:US-08-223-263-1 (1-353)
Number of sequences searched: 75511
Number of scores above cutoff: 4091
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       > 0 < O IntelliGenetics > 0 <
                                                                                                                                                                                                                                                                                                                                                                                                   F10000-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            U50000-+
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         Results file sqlpir.res made by
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         FastDB - Fast Pairwise Comparison of Sequences Release 5.4
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       100000-
                                                                                                                                                                                                                                                 1000-
                                                                                                                                                                                                                                                                                                                                                        5000-
                                                                                          100-
                                                                                                                                                                                                     500-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Results of the initial comparison of US-08-223-263-1 (1-353) with: Data bank : PIR 43, all entries
50---
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    TTHIANHSYL-PLS-620 630
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    X
dooma-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           on Tue 11 Apr 95 9:51:38-PDT.
```

Tue Apr.11 13:27:36 1995

Page 23

PARAMETERS

Similarity matrix Un Mismatch penalty Gap penalty Gap size penalty	Unitary 1 1.00 0.05	K-tuple Joining penalty Window size	20 32
Randomization group	01		
Initial scores to save Optimized scores to save	45 0	Alignments to save Display context	15 10
	SEAR	SEARCH STATISTICS	

Number of residues:	Times:	Scores:
2	CPU 00:02:33.97	Mean 6
22468834		Median 7
	Total Elapsed 00:02:34.00	Standard Deviation 2.60

Number of sequences searched: Number of scores above cutoff: 75511 4091

Cut-off raised to 5.
Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.
Cut-off raised to 10.
Cut-off raised to 11.
Cut-off raised to 11.

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

Sequence Name Description c-MPI ligand - human Init. Opt. Length Score Score Sig. Frame 353 353 353 133.36 0

The list of other best scores is:



Listing for Mary Hale

Tue Apr.11,13:27:36:1995

Page

43. A46132 44. S37679 45. PQ0475	•		•	•	•	•	. взэз	. S3289	. S261		GNWVT												19. A43325				FOMVM		13. FOMVMU		. 01101	10. 5244U/	. 54609	8. QYBYP	. S0/31	٠	. PQ045	. A2974		3. A37232	. of 0000) A	Sequence Name
un leucine z othetical pr til extensin	I protein - Xanthomonas	lagen alpha 2(IV) chain	- eseqtans down masord	protein - recvirus ty	protein - reovirus ty	protein - mouse	ic growth factor	otein 9 - Salm	ng ATP synthase	4 standard deviations	e polyprotein — hepatiti	iphysin - chicken	/protein - Moloney n	t protein - C	TV-XqT	beet	2	rich	ว	erycatopotecta precursor - mo	precursor -	precursor -	b-nydroxymethy	dihydro-6-	extensin-like protein precurs	Caldocel	in - Moloney mu	nosyltransferase (EC 2.	ב	rythronoietin - rat	deviations	roros	xylase	e carboxylase (EC 6.	ome-iron transport	ferrichrome-iron transport pr	otein - Per	epidermal growth factor-bindi	6 standard deviations	, tracheal (AMN-22) - hu	**** 7 standard deviations a	**** 72 standard deviations	Description
149 165 171	138	1707	180	462	399	389	350	299	248	bove mean	3010	682	538	υ . ω .	474	439	421	303	283	200	102	192	100	159	151	1331	538	480	468	192	3 8	1200	1180	1178	659	659	122	259	above mean	294	above mean	n	Length Sco
18 18	18	19	10	19	19	19	19	19	19		20	20	20	20	26	200	2	20	20	2 0) C	3 6	200	20	20	21	21	21	21	2		3 6	22	22	22	22	22	23	***		***	· *	re s
383	36	73	2 0	<u>ე</u> წ	63	26	60	43	36		89	59	ω G	60	5 :	40	χ 20	44	47	יי ני ני	ى د ئا د	2 0	ء ر <u>د</u>	ω	34	62	36	50	ω.	44	0/	30	2 6	68	59	59	30	46		55	27.2	n	core.
4.61 4.61							6				٠.	٠.			. د	. ن	٠		٠.	ن د	ن س	ن د	5.38 38	ů	ن	`.	7	7	. :	7			6.15	· ;_	÷	÷	÷	Ġ		7.69	12.23	ა ა	Sig. F
000	0	00	5 C	0	0	0	0	0	0		0	0	0	0 0	> (0 0	o	0 (0	> c	o c	o c	0	0	0	0	0	0	0	>	c	o C	0	0	0	0	0	0		0	c	•	Frame

1. US-08-223-263-1 (1-353) S45331 c-MPI ligand - human

ENTRY TITLE ORGANISM

\$45331 #type complete c-MPI ligand - human #formal_name Homo sapiens #common_name man



Tue Apr 77 13:27:36 1995

Page 25

Gaps Residue Identity Initial Score TITLE ORGANISM DATE SUMMARY SEQUENCE DATE ENTRY ACCESSIONS REFERENCE US-08-223-263-1 (1-353) S45330 thrombopoi MELTELLLIVMILLITARLTISSPAPPACDLRVISKLIRDSHVLHSRLSQCPEVHPLPTPVILPAVDFSIGEW
MELTELLIVMILLITARLTISSPAPPACDLRVLSKLIRDSHVLHSRLSQCPEVHPLPTPVILPAVDFSIGEW
MELTELLIVMILLITARLTISSPAPPACDLRVLSKLIRDSHVLHSRLSQCPEVHPLPTPVILPAVDFSIGEW
X
10
20
30
40
50
60
70
70
70 220 230 240 250 260 270 GLKWQQGFRAKIPGLLQTSRSLDQIPGYLNRIHELLNGTRGLEPGPSRRTLGAPDISGTSDTGSLPPNL GLLKWQQGFRAKIPGLLUQTSRSLDQIPGYLNRIHELLNGTRGLEPGPSRRTLGAPDISGTSDTGSLPPNL GLLKWQQGFRAKIPGLLQTSRSLDQIPGYLNRIHELLNGTRGLEPGPSRRTLGAPDISGTSDTGSLPPNL #journal #title #accession 220 ##residues ##status 230 thrombopoietin -0 0 0 10-Dec-1994 \$45331 \$45331 de Sauvage, F.J.; Hass, P.E.; Spencer, S.D.; Malloy, B.E.; Gurney, A.L.; Spencer, S.A.; Darbonne, W.C.; Henzel, W.J.; Wong, S.C.; Kuang, W.J.; Cles, K.J.; Hultgren, B.; Solberg Jr., L.A.; Goeddel, D.V.; Eaton, D.L. Nature (1994) 369:533-538
Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl ligand. \$45330 #type complete thrombopoietin - mouse #formal name Mus musculus #common name house 10-Dec-1994; #sequence_revision 10-Dec-1994; 10-Dec-1994 preliminary
1-353 ##label DEA
#length 353 #molecular-weight 37822 #checksum 1597 S45331 10-Dec-1994; #sequence_revision 10-Dec-1994; #text_change 90 353 100% Optimized Score Matches Conservative Substitutions 100 mouse 250 110 0 0 260 353 353 120 Significance Mismatches 270 130 #text_change 9 = 133.36 = 0 = 0 140

* ŵ

> ENTRY ORGANISM

A37232 #type fragment mucin, tracheal (AMN-22) - human (fragment) #formal name Homo sapiens #common name man 14-Feb-1992 #sequence_revision 14-Feb-1992

#text_change

1

ř

Listing for Mary Hale

Tue Apr 11 13:27:36 1995

26

murine thrombopoletin cDNA and production in vivo. production in vivo. = 252 Significance = 72.25	US-08-223-263-1 (1-353) A37232 mucin, tracheal (AMN-22) - human (fragment)	220 240 250 260 270 280 GLIKWQQGFRAKI-PGILIQTSRSIDQIPGYLNRIHELLAGTRGIFPGPSRRTIGAPDISGTSDTGSLFPN [130 11GALQSLLGTQLPPQG	Significance = 72. Mismatches = 8 SO 60 GCPEVHPLPTPVILPAVDFS	Blumberg, H.; Johnson, R.; Prunkard, D.; Ching, A.F.T.; Mathewes, S.L.; Bailey, M.C.; Forstrom, J.W.; Buddle, M.M.; Obborn, S.G.; Evans, S.J.; Sheppard, P.O.; Presanell, S.R.; #journal Nature (1994) 395:565-568 #title Cloning and expression of murine thrombopoietin cDNA and stimulation of platelet production in vivo. #accession S4530 ##residues 1-356 ##alabel LOK SUMMARY #length 356 #molecular-weight 37835 #checksum 9983 SEQUENCE	REFERENCE \$45330 #authors Iok, S.; Kaushansky, K.; Holly, R.D.; Kuijper, J.L.; Lofton-Day, C.E.; Oort, P.J.; Grant, F.J.; Heipel, M.D. Burkhead, S.K.; Kramer, J.M.; Bell, L.A.; Sprecher, C.A
---	---	---	-----------------------	---	---	--

* Tue Apr 11 13:27:36 1995

27

Listing for Mary Hale

Tue Apr 11

Page

4. US-08-223-263-1 (1-353 A29746 epidermal ORGANISM DATE Residue Identity Initial Score ENTRY ACCESSIONS REFERENCE SEQUENCE **DOMMARY** TITLE 30-Sep-1993
CERSIONS A37232
FERENCE A37232

#authors Gerard, C.; Eddy Jr., R.L.; Shows, T.B.
#journal J. Clin. Invest. (1990) 86:1921-1927

#title Contains a tandem repeat structure. Evidence for a common mucin in airway and gastrointestinal tissue.
#cross-references MUID:91072667
#accession A37232 SLLGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAP-----TSTQTPTPTPITT LINGTRGLFPGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPTLPTPVVQLHP--200 210 220 230 240 LIBELPIRTSGLLET-NFTASARTTGSGLLKWQQGFRAKIPGLLNQT----SRSLDQIP----LPTPVLLPAVDFSLGEWKTQ-MEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQ ---LLPDP-##cross-references GB:M57417
Y #length 294 #checksum 1596 ##molecule_type mRNA ##residues ##status 260 190 TPTPTGTQTPTPTPTTTTTMVT-X 10 20 0 0 0 TVTPTPTPTSTØRTTPTSITTTTVTPTPT #formal name Mus musculus #common_name house mouse
08-Mar-1989 #sequence_revision 30-Jun-1989 #text_change A29746 #type complete epidermal growth factor-binding protein type A precursor mouse 8 26 21**%** 61 -ÞTSÍ-----TTTTTVTPTPTPITTTTTVTPTPTBTGTQTPTTTBISTTTTV-T 60 70 80 90 100 1-294 ##label GER preliminary growth factor-binding taa Optimized Score Matches Conservative Substitutions RLLNTSYTHSQNLSQEG 340 160 290 350 220 170 ----PTFTITSTQTPTPTTTTTTVTPTP protein type A pre 55 70 Significance Mismatches 30 102(6) 180 PTTAVPSRTSLVLT 0 0 0

-GYLNRIHE Gaps Residue Identity Initial Score ENTRY TITLE CLASSIFICATION FEATURE ACCESSIONS REFERENCE ORGANISM SUMMARY ACCESSIONS SEQUENCE US-08-223-263-1 (1-353) PQ0452 KNSQPWQVAVYYLDEYLCGGVLLDRNWVLTA--40 X 50 60 200 210 220 230 240 LIVELPHRTSGLIETHFTASA-RTTGSGLIKWQQGFRAKIPGLINQTSRSL NPNDLQCVSIKLHPNEVCVKAHILKVTDVMLCAGEMNGGKDTCKGDSGGPLIC-170 180 190 200 210 130 140 150 160 170 180 190 LIGALQSILIGTQLPPQGETTAH-KDENAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPPTTAVPSRTSLVLJ #authors Drinkwater, C.C.; Evans, B.A.; Richards, R.I.
#journal Biochemistry (1987) 26:6750-6756
#title Mouse glandular kallikrein genes: identification characterization of the genes encoding the epid factor binding proteins.

#cross-references MUID:88107594 HPDENMSLLQSVPTGADLSNDLMLLRLSKPADLTDVV--KPIDLPTTEPKLGSTCLASGWGSINQLIYQ---HPLPTPVLLPAV--DFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLG----QLSGQVRL 25-251 #accession *contents ##residues ##molecule_type DNA 8 X 10 20 30 MELTELLLVVMIL-----LTARLTLSSPAPPACD--LRVLSK--LLRDSHVLHSRLSQCPEV extensin-like protein - Persian tobacco (strain H H H 28-Oct-1994 A29746 PQ0452 #type fragment
extensin-like protein - Persian tobacco (strain S2S3)
(fragment) #superfamily trypsin; trypsin homology A29746 #formal name Nicotiana alata #common name Persian tobacco 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993 #domain trypsin homology #label TRP #length 259 #molecular-weight 28384 #checksum Strain BALB/c, salivary gland A90522 APAIYTKLIKFTSWIKDTMAKNP 23 23 45 1-259 ##label DRI
the authors translated the codon Conservative Substitutions Optimized Score Matches 80 90 -AHCYEDKYNIWLGKNKLFQDEPSAQHRLVSKSFP 70 80 90 H H 100 46 59 Significance = Mismatches = for tion and epidermal -DGVLQGITSWGST 220 230

120



¥

N. A.

28

residue

210

as

growth

2468

6.53 142 0

13:27:36 1995

Tue Apr 11 13:27:36 1995

Page 29

Initial Score Residue Identity Residue Identity Initial Score SUMMARY SEQUENCE ENTRY TITLE ORGANISM ACCESSIONS REFERENCE ##residues SUMMARY SEQUENCE REFERENCE US-08-223-263-1 (1-353)
S45222 ferrichrome-iron transport protein -RDAHGWTWASGALLEDLMPWRWPRIMAALFAGVML---AVAGCIIQRLTGNPMASPEVLGISSGAAFGVVIM 370 380 390 400 410 420 430 300 310 320 330 340 350 X
GYSPSPTHPPTGQYTLFPLFPTLPTPVVQLHPLLPDPSAPTPTTSPLLNTSYTHSONLSQEG #journal #title PPSPSPPPPSPSPPPPSPSPPPPPSPPADDMAPSPSP-AAAPTPPDMLLRPSARLAGNMVISGLH #accession #submission #authors FRAKIPGLINGTSRSLDQIPGYLNRIHELLNGTRGLFP--#accession *contents fauthors ##molecule type mRNA ##residues 1-122 ##label CHE ## #hength 122 #checksum ##status preliminary ##residues 1-59 ##label FUJ ##cross-references EMBLID26562 ##cross-references EMBLID26562 Y #length 659 #molecular-weight 70335 HGTESL--PDATLQR-FKFKWPFFGKSPKNSPKSSPSRSPPPKREQPSPPPPVKSPPPPSPP X 10 20 30 40 50 MELTELL-----LVVMLLLTARLTLSSPAPPACDLRVLSKLLRDS-HVL-HSRLSQCPEVHP H 0 H II II II Specific expression of an extensin-like gene in the style of Nicotiana alata.
Style, Strain S2S3
PQ0452 ferrichrome-iron transport protein - Escherichia coli
#formal name Escherichia coli
07-0ct-1994; #sequence_revision 07-0ct-1994; #text_change S45222 S45222 submitted to the EMBL Data Library, January 1994 Fujita, N. \$45181 07-Oct-1994 22 24% 77 22 28**%** 11 80 Optimized Score = 30 Matches = 35 Conservative Substitutions Matches = 79 Conservative Substitutions Optimized Score #type complete 260 208 11 11 100 ----GPSRRTLGAPDISSGTSDTGSLPPNLQP 59 79 Escherichia c Mismatches Mismatches Significance = Significance = #checksum 1419 40 н 6.15 170 0 6.15 79 0

Listing for Mary Hale

Tue Apr 11 13:27:36 1995

> Page ဗ

```
#map_position (
#start_codon (
CLASSIFICATION #
KEYWORDS
FEATURE
1-20
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     7. US-08-223-263-1 (1-353)
S07318 ferrichrom
                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ACCESSIONS
REFERENCE
#authors
#journal
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ORGANISM
DATE
  Initial Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ENTRY
TITLE
                                                          SUMMARY
                                                                                                                                                                                                                                                       GENETICS
                                        SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               MALTPTRIAL--
560
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    180 190 200 210 220 230 240 250 TTAVPSRTSLVITLNELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRI
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             SALVGGLLLVFADWCGRMVLFPFQIPAGLLSTFIGAPYFIYLLRKQSR
620 630 640 650 X
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               HELLNGTRGLF-PGPSRRTLGAPDISSGTSDTGSLPP---NLQPGYSPSPTHPPTGQY
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                LLMMLQ-ASGDPRMAQVLTWISGSTYNATDAQVWRTGIVMVILLAITPLCRRWLTILLPL-GGDT---ARAVG 500 510 520 530 540 550
                                                                                                21 - 659
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        LLGALQSLLGTQLPPQ-----
                                                                                                                                                                                                                                 #gene
                                                                                                                                                                                                                                                                                                                                                                     #accession
                                                                                                                                                                                                                                                                                                                                                                                                                                                    #title
                                                                                                                                                                                                                                                                                                                                                                   cross-references MUID:87014116
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  --FLVPGNAFGWLLPAGSLG-440 450
                                                                                                                                                                                                                                                                                                                                 ##molecule_type DNA
##residues 1-65
                                                                                                                                                                                                                                                                                                           ##cross-references EMBL:X04319
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 ferrichrome-iron transport protein fhuB precursor
    ti
                                                                                                                                                                                                              fhuB
4 min
                                                                                                                                                                                            GIG
                                                                                                                                                                                                                                                                                                                                                                                                                     Koester, W.; Braun, V.
Mol. Gen. Genet. (1986) 204:435-442
Iron hydroxamate transport of Escherichia coli: nu
sequence of the fhuB gene and identification of
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ferrichrome-iron transport protein fhuB precursor -
Escherichia coli
#formal name Escherichia coli
#formal name Escherichia coli
31-Dec-1900 #sequence revision 31-Dec-1990 #text change
                                                        #length
                                                                                                                                                                    #superfamily vitamin B12 transport protein btuC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            S07318
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            S07318
                                                                                                                                                                                                                                                                                                                                                                                                         protein.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              18-Jun-1993
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           270
                                                                                                                                                    transport; membrane protein
                                                      #domain signal sequence #status predicted #label SIG\
#product ferrichrome-iron transport protein fhuB #status
predicted #label MAT

jth 659 #molecular-weight 70335 #checksum 1419
                                                                                                                                                                                                                                                                         the authors translated the codon ATC for residue 311 as {\tt Thr}
                                                                                                                                                                                                                                                                                                                               1-659 ##label KOE
  22
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             ---LLIAACLTATATMTIGPL--SFVGLMAPHIARMMGFRRTMPHI-
570 580 590 600
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          -GRTTAHKD--
Optimized Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          #type complete
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                150 160 170
--PNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPP
    ()
  59
Significance =
                                                                                                                                                                                                                                                                                                                                                                                                                                                 nucleotide
  6.15
```

Tue Apr 11-13:27:36 1995

Page 31

ENTRY ENTRY Pyruvate portage ALTERNATE NAMES pyruvic ORGANISM	260 HELLNGTRGLF-PGH	130 LIGALOSLICTOLPPO	60 TPVLLPAVDF- FLVPGNAFG	X MELTE HELTE RDAHGWTWASGALLE 370 380	Residue Identity = Gaps =
PRINATE NAMES pyruvate carboxylase (EC 6.4.1.1) 1 - yeast (Saccharomyces RNATE NAMES pyruvic carboxylase (EC 6.4.1.1) 1 - yeast (Saccharomyces NISM	260 270 280 290 X 300 HELLINGTRGLF-PGPSRRTIGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQY	130 130 130 130 130 140 150 160 170 144 114 115 116 117 117 117 117 117 117 117 117 117	GWLLPAG	10 20 30 40 50 LTELLLVVMLLLTARLTLSSPAPPACDLRVLSKLLRDS-HVL-HSRLSQC	

Listing for Mary Hale

"Tue Apr 11 13:27:36 1995

```
Initial Score = Residue Identity = Gaps =
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      SUMMARY
SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   FEATURE
157-331
353-468
569-908
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         #map_position
CLASSIFICATION
KEYWORDS
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         GENETICS
                                                                                                                                                                                                                                                                           PNIPFQMÍLRGANGVAYSSIPDNAIDHFVKQÁKDNSVDIFRVFDAINDLEQLKVGVDAVKKAGGVVEATVCF
630 640 650 660 670 680 690 700
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         ##molecule_type DNA
##residues 103-1178 ##label MOR
##cross-references EMBL:J03889
#accession A29722
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                DTTWRDAHQSLLATRVRTHDLATIAPTTAHALAGAFALECWGGATFDVAMRFLHEDPWQRLRKLR----SLV 570 580 590 600 610 620
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             RIRGVKTNIPFLLTILTNPVFIEGTYWGT-FIDDTPQLFQMVSSQNRAQKLLHYLADVADNGSSIKGQIGLP
440 X 450 460 470 480 490 500
-HDSAGTRVASMTACALAGADVVDVAINSMSGL--TSQPSINALLASLEGNIDTGINVEH
770 780 790 800 810 X 820
                                                         310 320 330 340 350 X
LFPLEPTLPTEVVQLHPLLPPSAPTETEFLINTSYTHSQNL--SQEG
                                                                                                                                   SGDMLQPGKKYNLDYYLEIAEKIVQMG--THILGIKD-MAGTMKPAAAKLLIGSLRAKYPDLPIHVHT----710 720 730 740 750 760
                                                                                                                                                                                                 230 230 PSRTSLVLTLNELENRTSGL---LETNFTASA-----RTTGSGLLKWQ--QGFRA--KIPGLLNQT-SR
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           60 70 110 120
TPVILPAVDFSLGEMKTOMEETKAQDILGAVTLLLEGVMA---REGOLGPTCLSSLLGQLSGQVR-----LL-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           1089-1178
                                                                                                                                                                                                                                                                                                                                                                                                                                                              ----LGALQSLLGT-----QLPPQGRTTAHKDPNAIFISFQHLLRGKV--RFIMIVGGSTLCVRRAPPTTAV
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 gene
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              ##molecule type protein
##residues 1124-1149 ##label MOR2
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             MELTELLLVVMLLLTARLTLSSPAPPACDLRVLS----KLLRDSHVLHSRLSQCPEVHPLP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              LISTA:PYC1
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            #domain ATP/bicarbonate-binding #label ATB1\
#domain ATP/bicarbonate-binding #label ATB2\
#domain pyruvate-binding #label PYR\
#domain biotinyl- or lipoyl-binding #label BIO\
#binding site biotin (Lys) (covalent) #status predicted
#length 1178 #molecular-weight 130098 #checksum 6631
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         #superfamily pyruvate carboxylase
biotin; gluconeogenesis; homotetramer; ligase; zinc
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      Optimized Score = Matches =
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        68 Significance = 92 Mismatches =
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    --QVRQFNGTLLM
550 560
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     11 12
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  6.15
229
0
```



9. US-08-223-263-1 (1-353)

pyruvate carboxylase (EC 6.4.1.1) 2 - yeast (Sacch

Initial Score Residue Identity #map position a 1136 SUMMARY SEQUENCE ENTRY TITLE REFERENCE #authors GENETICS REFERENCE ACCESSIONS ORGANISM ALTERNATE_NAMES KIKSNPSVPHLHDAQGNVINVTKSAPPSGWRQVLLE-510 520 530 540 RIRGVKTNIPFLLTLLTNPVFIEGT-YWTTFIDDTPQLFQMVSSQNRAQKLLHYLADLAVNGSSIKGQIGLP #authors Stucka, R.; Dequin, S.; Salmon, J.M.; Gancedo, C.
#journal Mol. Gen. Genet. (1991) 229:307-315

#title DNA sequences in chromosomes II and VII code for pyruvate carboxylase isoenzymes in Saccharomyces corevisiae:
analysis of pyruvate carboxylase-deficient strains.
#cross-references MUID:92017667 #accession #submission TPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLL-----LGA faccession 8 ##molecule_type DNA #residues 1-1180 ##label DUB ##cross-references EMBL:Z36087 ##cross-references EMBL:X59890 X 10 20 30 40 50 MELTELLIVVMLLITARLITISSPAPPACDIRVLS-----KLLRDSHVLHSRLSQCPEVHPLP 0 11 11 protein YBR1507; protein YBR218c #formal name Saccharomyces cerevisiae 26-Aug-T994 #sequence_revision 09-Sep-1994 #text_change 14-Sep-1994 \$46094; \$17469 Dubois, E.; El Bakkoury, M.; Glansdorff, N.; Messenguy, Pierard, A.; Scherens, B.; Vierendeels, F. submitted to the Protein Sequence Database, August 1994 #superfamily pyruvate carboxylase
biotin; gluconeogenesis; homotetramer; ligase; zinc s46094 #type complete pyruvate carboxylase (EC 6.4.1.1) S46094 S17469 #length 1180 LISTA:PYC2 23 23 65 80 Optimized Score Matches Conservative Substitutions 11 11 ----KGPSEFAKQVRQFNG--TLLMDTTWRDA 91 91 2 Significance Mismatches yeast (Saccharomyces H 6.15 237 0 H



Listing for Mary Hale

Tue Apr 11 13:27:37 1998

Page

10. US-08-223-263-1 (1-353) S24407 formin isc formin isoform IV - mouse

ACCESSIONS
REFERENCE
#authors
#journal DATE ENTRY SUMMARY ORGANISM #authors Jackson-Grusby, L.; Kuo, A.; Leder, P.
#journal Genes Dev. (1992) 6:29-37
#title A variant limb deformity transcript expressed in the
embryonic mouse limb defines a novel formin.
#cross-references MUID:92112033 #accession ##status ##cross-references EMBL:X62379
#length 1206 #molecular-weight 133463 #checksum ##residues formin isoform IV - mouse
#formal name Mus musculus #common name house mouse
19-Feb-1994; #sequence_revision 19-Feb-1994; #text_change S24407 19-Feb-1994 preliminary 1-1206 ##label JAC #type complete 501

Residue Identity [nitial

H H H

Optimized Score Matches

67 95

Significance = Mismatches =

6.15 220 0

Conservative Substitutions

Score



.

Page ઝ

11. US-08-223-263-1 (1-353) S11515 formin - m formin - mouse

ENTRY
TITLE
ORGANISM
DATE #formal name Mus musculus #common name house mouse 22-Jan-1994; #sequence_revision 22-Jan-1994; #text_change 22-Jan-1994 S11515 formin - mouse #type complete

ACCESSIONS REFERENCE #authors #journal Woychik, R.P.; Maas, R.L.; Zeller, R.; Vogt, T.F.; Leder, P. Nature (1990) 346:850-853
'Formins': proteins deduced from the alternative transcripts of the limb deformity gene.

#accession S11515

SUMMARY SEQUENCE ##status preliminary ##residues 1-1468 ##label WOY ##cross-references EMBLY.53599 ##Cross-references EMBLY.53599

Residue Identity Initial Score 22 23* 86 Conservative Substitutions Matches Optimized Score 11 11 67 95 40 Mismatches Significance H II 6.15 224 0

Listing for Mary Hale

Tue Apr. 11:13:27:37 1995

36

Residue Identity ENTRY TITLE ORGANISM 12. US-08-223-263-1 (1-353) S28148 erythropoi Initial Score DATE SEQUENCE REFERENCE ACCESSIONS 250 MGVPERPTILILLSLILIPUGLPVLCAPPRLICDSRVLERYILEAKEAENVTMGCAEGPRLSENITVPDTKV X 10 20 30 40 50 60 70 PKKLTISLTQLSPSKDSKDIHAPFQTREGTSSSSQQKISPPAPPTPPPLIPPPLIPPPPPPPG---LGPLPP 840 850 860 870 880 890 #accession ##status TLPTPVV--QLHPLLPDPSAPTPTPTS--EEAIENIKQOLEKRREGCEEMRDVCISTDDDCSPKAFRNVCIQTDRETFIKPCDAESKATRS-SQI---770 780 790 800 810 820 WPPPKTKDTEEKVGLKYTEAEYQAAI----LHLKREHKEEIETLQAQFELKTFHIRGEHALVTAR-----L 710 720 730 740 750 760 TAKDLPNKD--GGVWVPGYRAGPPCPFL----LHEEKEKTSRSEL---YLDLNPDQSPTEQDDRTPGRLQAV IHELLNGTRGLFPGPSRRTLGAP-DISSGTS----DTGSLP----PNLQPGYSPSPTHPPTGQYTLFPLPP #journal #authors TPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQ--LSGQVRLLLGALQ--130 140 150 160 170 180 190 ----SILGTQLPPQGRTTAHKDPNAIFLSFQHLLR-GKVRFIMLVGGSTLCVRRAPPTTA-VPSRTSLVLTL title MELTELLLVVMLLLTARLTLSSPAPP---ACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTPVLLPAVDF ##residues 200 210 220 230 240 -- EURITSGLLE-TNFTASARTTGS---GLLKWQQGFR---AKIPGLLNQTSRSLDQIPGYLNR H H erythropoietin -Nagao, M.; Suga, H.; Okano, M.; Masuda, S.; Ikura, K.; Sasaki, R. Biochim. Biophys. Acta (1992) 1171:99-102 \$28148 #type complete
erythropoietin - rat
#formal name Rattus norvegicus #common name Norway rat
22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change #length 192 #molecular-weight 21286 #checksum 4587 Nucleotide sequence of rat erythropoietin. S2814 S28148 S28148 22-Nov-1993 21 27**%** 24 preliminary 1-192 ##label NAG Optimized Score Matches Conservative Substitutions 280 -PLLNTSYTHSQNLSQEG 11 19 44 53 350 × Significance = Mismatches = 50 Narita, H.; 11 11 5.77 118 0

100

.00 110 120 -RGQLGPTCLSSLLGQLSGQVRLLLGA---

Listing for Mary Hale

Tue Apr 11 13:27:37 1995

38

```
#gene
CLASSIFICATION
KEYWORDS
                                                                                                                                   Residue Identity
                                                                                                                                                   Initial Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              13. US-08-223-263-1 (1-353)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   ENTRY
TITLE
                                                                                                                                                                                   SEQUENCE
                                                                                                                                                                                                   SUMMARY
                                                                                                                                                                                                                                                                                           FEATURE
                                                                                                                                                                                                                                                                                                                                                        GENETICS
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             REFERENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             ACCESSIONS
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         ORGANISM
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      ALTERNATE_NAMES
CONTAINS
                                                                 170 180 190 200 210 220 220 TICVRRAPPTTAVPSRTSIVLTINELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLD
                                                                                                                                                                                                                                                                                                                                                                                                                                                 #journal
#title
                                                                                                                                                                                                                                216-468
                                                                                                                                                                                                                                                             132-215
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            LMSPPDATQAAP-LRTLTADTFCKLFRVYSNFLRGK---LKLYTG-EACRRGDR 150 160 170 180 190 X
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        {\tt LQS-LLGTQLPPQGRTTAHKDPNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPPTTAVPSRTS}
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               #authors
                                                                                                                                                                                                                                                                                                                                                                                       ##residues
                                                                                                                                                                                                                                                                                                                                                                       250
                              MGQTVTTPLSLTL-
X 10
                                                                                                                         . . .
                                                                                                                                                                                                                                                                                                                                                                                                               Huai, I.; Chiocca, S.M.; Gilbreth, M.A.; Ainsworth, J.R.; Bishop, L.A.; Murphy Jr., E.C. J. Virol. (1992) 66:3329-5337
Moloney murine sarcoma virus MuSVts110 DNA: cloning, nucleotide sequence, and gene expression.
A42745
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               gag polyprotein - Moloney murine sarcoma virus (st
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 FOMVMU #type complete
gag polyprotein - Moloney murine sarcoma virus (strain
MuSVts110)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           140
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    core polyprotein
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      core protein p15; core shell protein p30; inner coat protein
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       #formal_name Moloney murine sarcoma virus
30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
                                                                                                                                                                                                  #length 468  #molecular-weight 52681  #checksum 3522
                                                                                                                                                                                                                                                                                                                        superfamily mammalian retrovirus gag polyprotein I
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          08-Apr-1994
260
                                                                                                                        21
22%
15
                                                                                                                                                                                                                                                                                                        protein; core protein; polyprotein
                                                                                                                                                                                                                   Conservative Substitutions
                                                                                                                                      Optimized Score
Matches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           150
270
                                            -DHWKDVERIAHNQSVDVKKRRWVTFCSAEWP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           160
                                                                                                                                         H 11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           170
                                                                                                                                      35
40
 290
                                                                                                                                      Mismatches
                                                                                                                                                   Significance
 300
 310
                                              TFNVGW
                                                                                                                        120
```

14. US-08-223-263-1 (1-353) #gene KEYWORDS FEATURE 16-30 Residue Identity Initial Score SUMMARY SEQUENCE ACCESSIONS REFERENCE ENTRY GENETICS ORGANISM STPPQSSLYPAL-TPS-LGAKPKPQVLSDSGGPLIDLLTEDPPPYRDPRPP 130 140 150 160 170 320 330 X PT-PVVQLHPLLPDPSAETPTPTSPLLNTSYTHSQNLSQEG QIPGYLNRIHELLNGTRGLFPGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLPPTL 90 100 110 120 130 140 150 LGAVILLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQSLLGTQLPPQGRTTAHKDPNAIFLSFQHLL TLSSPAPPACDLEVLSKL--LEDS--HVLHSRLSQCPEVHPLPTPVLLPAVDFSLGEWKTQMEET--KAQDI 203, 281, 341, 393 PRDGTFNRDLITQVKIKVFSPGPHGHPDQVPYIVTWEALAFDPPPWVKPFVHPKPPPPLLPSAPSLPLEPPL ---NLKKQITVNKKKNQ--#journal EMBO J. (1992) 11:2511-2519
#title OCH1 encodes a novel membrane bound mannosyltransferase:
 outer chain elongation of asparagine-linked
 oligosaccharides.
#cross-references MUID:92331603
#accession S22701 #authors ##molecule_type DNA ##residues 1-48 ##cross-references EMBL:D11095 MSRKLSHLIATRKSKTIVVTVLLIYSLLTFHLSNKRLLSQFYPSKDDFKQTLLPTTSHSQDIX 10 20 30 40 50 60 0 0 0 mannosyltransferase (EC 2.4.1.-) - yeast (Saccharo #domain transmembrane #status predicted #label TMM\
#binding site carbohydrate (Asn) (covalent) #status
predicted
predicted
#length 480 #molecular-weight 55155 #checksum 9502 LISTA:OCH1 glycoprotein; transmembrane protein S22701 #type complete mannosyltransferase (EC 2.4.1.-) -Nakayama, K.I.; Nagasu, Jigami, Y. #formal_name Saccharomyces cerevisiae 04_Dec-1992 #sequence_revision 04-Dec-1992 #text_change Jigami, 09-Sep-1994 cerevisiae) 21 20% 43 1-480 ##label NAK --LHNLRDQLSFAFFYDSQAPIPQRVWQTWKVG---ADDKNFPSSFRTYQKTW 80 90 100 110 120 40 Optimized Score = 56 Matches = 73 Conservative Substitutions Τ.; Shimma, Significance = Mismatches = = yeast ۲.; Kuromitsu, (Saccharomyces ٠. :: 5.77 236 0



127

.

15. US-08-223-263-1 (1-353) FOMVM gag polypr

gag polyprotein - Moloney murine sarcoma virus

ACCESSIONS REFERENCE ENTRY TITLE ORGANISM CONTAINS #authors Reddy, E.P.; Smith, M.J.; Aaronson, S.A.
#journal Science (1981) 214:445-450
#title Complete nucleotide sequence and organization of the Moloney
murine sarcoma virus genome.
#cross-references MUID:82039559 FOMVM #type complete
gag polyprotein - Moloney murine sarcoma virus
core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein p10
#formal name Moloney murine sarcoma virus
18-Dec-1981 #sequence_revision 18-Dec-1981 #text_change
30-Sep-1933 A94261; B00645; A03932 protein

A94261 Provirus

#accession

##residues #molecule_type genomic RNA #residues 1-538 ##label RED

REFERENCE #title Nucleotide sequence of the genome of a murine sarcoma virus. #cross-references MUID:82115347 #contents Clone 124, circular #journal fauthors Cell (1981) 27:97-108 Van A00645 I.M. Beveren, C.; van Straaten, F.; Galleshaw, J.A.; Verma,

B00645

##molecule type DNA
##molecule type DNA
1-518,'K',520-538 ##label VAN

GENETICS #gene

CLASSIFICATION FEATURE gag #superfamily mammalian retrovirus gag polyprotein I

Listing for Mary Hale

Tue Apr.11 13:27:37 1995

6

Gaps Residue Identity FastDB - Fast Pairwise Comparison of Sequences Release 5.4 Initial Score SUMMARY SEQUENCE 132-215 216-478 479-534 VGWPRDGTFNRDLITQVKIKVFSPGPHGHPDQVPYIVTWEALAFDPPPWVKPFVHPKPPPPLLPSAPSLPLE
50 60 70 80 90 100 110 320 330 340 350 x ${\tt SLDQIPGYLNRIHELLNGTRGLFPGPSRRTLGAPDISSGTSDTGSLPPNLQPGYSPSPTHPPTGQYTLFPLP}$ ${\tt TLCVRRAPPTTAVPSRTSLVLTIMELPNRTSGLLETNFTASARTTGSGLLK----WQQGFRAKIPGLLNQTSR}$ IntelliGenetics 180 MGQTVTTPLSLTL--DHWKDVERLAHN--X H 11 H #product core protein p15 #label P15\
#product inner coat protein p12 #label P12\
#product core shell protein p30 #label P30\
#product nucleoprotein p10 #label P10
#length 538 #molecular-weight 61209 #checksum ! 21 24% 21 260 Optimized Score = 36 Matches = 43 Conservative Substitutions -QSVDVKKRRWVTFCSAEWP----TFN 30 290 Significance = Mismatches = #checksum 590 5.77 114 0

Results file sqlspt.res made by on Tue 11 Apr 95 9:56:26-PDT.

Query sequence being compared:US-08-223-263-1
Number of sequences searched:
Number of scores above cutoff: 1 (1-353) 40292 4111

Results of the initial comparison of US-08-223-263-1 (1-353) with: Data bank : Swiss-Prot 30, all entries

10000-1000-5000-500-



Tue Apr 11 13:27:37 1995

Page 41

O---11
SCORE 0|
STDEV -2 **೧೯೧**೫೯೦ ¹-ω= œ ---414 17 σ--19 8-22-25-

PARAMETERS

Scores:		Initial scores to save Optimized scores to save	Similarity matrix Mismatch penalty Gap penalty Gap size penalty Cutoff score Randomization group
Mean 6	SEAR	e 45	Unitary 1.00 0.05 2
Median 8	SEARCH STATISTICS	Alignments to save Display context	K-tuple Joining penalty Window size
Standard 2.07	S	to save ntext	nalty . e
Standard Deviation 2.07		15 10	32 32

Number of residues: Number of sequences searched: Number of scores above cutoff: 14147368 40292 4111 Times:

CPU 00:01:20.11

Total Elapsed 00:01:22.00

Cut-off raised to 6.
Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.
Cut-off raised to 10.

Listing for Mary Hale

Tue Apr 11 13:27:37 1995

Page 42

Cut-off raised to 11.

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

30 32 32 33 33 34 35 36 37 37 40 40 42.	116 117 118 119 119 119 119 119 119 119 119 119	11111111111111111111111111111111111111	Sequence
GSPI XANCP YATZ SYNP6 DEVS MYXXA EXTN DAUCA SPYA CALJA EBNZ EBV CPSI MOUSE NU62 HUMAN DCPY HANUV V70K PLRVW CNA2 FILCA DPO1 BACCA	HPPK EXTN EXTN ACRO VTP3 HUP1 GAG N POLG ATPO UGG1 VS11 VS11 VS2 ARM N	EPO_FELCA EPO_CANFA KLKI_MOUSE FHUB_ECOLI PYC1_YEAST PYC2_YEAST PYC2_YEAST PYC2_YEAST PYC2_YEAST PORM_MOUSE EPO_RAT EPO_MOUSE EPO_RAT EPO_MOUSE GAG_MIVAB GAG_MSVMT GCHI_YEAST GAG_MSVMT G	nce Name
BECRETION PATHWAY PRO HASE SUBUNITS REGION BODY DEVELOPMENTAL I PRECURSOR. PRECURSOR. PRECURSOR AMINOTRANSFI ICLEAR PROTEIN (EC 1.1) POPE GLYCOPROTEIN P66. DECARBOXYLASE (EC 4.1) PROTEIN (ORF 2). PROTEIN (ORF 2). PROTEIN (ORF 2). PROTEIN (ORF 2). PROTEIN (ORF 2). PROTEIN (ORF 2).	deviations cursor. (PROLINE-I (EC 3.4.21 (EC	**** 9 standard deviations a ERYTHROPOIETIN PEECURSOR (FFA **** 8 standard deviations a EXTHROPOIETIN PRECURSOR (FFA GIANDULAR KAILIKEEIN KZ2 PREC **** 7 standard deviations a FERRICHROME TRANSPORT PROTEIN PYRUVATE CARBOXYLASE 1 (EC 6. FYRUVATE CARBOXYLASE 2 (EC 6. FYRUVATE CARBOXYLASE 2 (EC 6. FORMIN 4. FORMIN PRECURSOR. ERYTHROPOIETIN PRECURSOR. ERYTHROPOIETIN PRECURSOR. ERYTHROPOIETIN PRECURSOR. GAG POLYPROTEIN (CONTAINS: CO GAG POLYPROTEIN (CONTAINS: CO GAG POLYPROTEIN RES (CONTAINS EAGR	Description
138 208 214 306 414 487 487 522 564 639 686 687	above mean 1158 1192 283 283 283 301 474 474 301 533 538 538 530 546 246 469 469 8169 81707	above mean 188 above mean 175 259 above mean 1180 659 alone 1180 1180 1192 192 235 235 235 235 338 1331	Init. Length Score
18 118 118 118 118 118 118 118 118 118	**************************************	* * * * * * * * * * * * * * * * * * *	
36 43 42 58 58 33 33 40 62 62 62	33 443 446 60 61 60 60 60 60 60 60 60 60 60 60 60 60 60		re·
5.		(R(R(R(R(R(R(R(R(R(R(R(R(R(R(R(R(R(R(R	Sig. Fr
00000000000	000000000000	000000000000000000000000000000000000000	Frame

43. 44.

POLG YEFV2
POLG YEFV1
AMEG_BOVIN

GENOME POLYPROTEIN GENOME POLYPROTEIN AMELOGENIN.

(CONTAINS:

3411 3411 170

18 18 17

48 48 36

5.80 5.32

000

US-08-223-263-1 (1-353) EPO_FELCA ERYTHROPOIETIN

PRECURSOR

(FRAGMENT).

Tue, Apr 11 13:27:37 1995

Page

43

```
Residue
                                                                                                                                                                   Initial
                                                                                                                                                                                         MELTELLLVVMLLLTARLTLSSP---APP--ACDLRVLSKLLRDSHVLHSRLSQCPEVHPLPTFVLLPAVDF
140
                   NFYTWKRMDVGQQAVEVWQGIAILSE—AILRGQALLANSSQPSETIQLHVDKAVSSIRSITSLIRAIGAQ
70 80 90 100 110 120 130
                                                 70 80 90 100 110 120 1 SIGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQLSGQVRLLLGALQSL--
                                                                                 ECPALLILLSLLI---LPIGLPVLGAPPRLICDSRVLERYILEAREAENVTMGCAEGCSFSENITVPDTKV X 10 20 30 40 50 60
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    01-FEB-1994
01-FEB-1994
01-FEB-1994
                                                                                                                                                                                       CARBOHYD
CARBOHYD
CARBOHYD
SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   EPO FELCA P33708;
                                                                                                                                                                                                                                 DISULFID
                                                                                                                                                                                                                                                                                                                             GOODMAN M., BUNN H.F.;

BLOOD 82:1507-1516(1993).

11-00-82:1507-1516(1993).

11-01-FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.

11-TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MANMA AND BY LIVER OF FETAL OR NEONATAL MAMMALS.
                                                                                                                                                                                                                                                      CHAIN
                                                                                                                                                                                                                                                                 SIGNAL
                                                                                                                                                                                                                                                                           NON TER
                                                                                                                                                                                                                                                                                                          EMBL;
                                                                                                                                                                                                                                                                                                                                                                                                       WEN D.,
                                                                                                                                                                                                                                                                                                                                                                                                                 93372347
                                                                                                                                                                                                                                                                                                                                                                                                                            SEQUENCE FROM N.A.
                                                                                                                                                                                                                                                                                                                                                                                                                                               EUTHERIA; CARNIVORA
                                                                                                                                                                                                                                                                                                                                                                                                                                                         FELIS CATUS (CAT).
EUKARYOTA; METAZOA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 EPO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ERYTHROPOIETIN
                                                                                                                                                                                                                                                                                     ERYTHROCYTE MATURATION;
                                                                                                                                                                                                                                                                                                PROSITE;
                                                                                                                                                      Identity
                                                                                                                                                                     Score
                                                                                                                                                                                                                                                                                                                     SUBCELLULAR LOCATION: SECRETED
                                                                                                                                                                                                                                                                                                          L10606; FCERYTHRO.
                                                                                                                                                                                                                                                                                                                                                                                                      BOISSEL J.P.R.,
150
                                                                                                                                                 11 11 11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     (REL.
(REL.
                                                                                                                                                                                         188
                                                                                                                                                                                                  1
<1
23
29
51
46
60
                                                                                                                                                                                        ₽,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          PRECURSOR (FRAGMENT).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               STANDARD;
                                                                                                                                              25
25*
26
                                                                                                                                                                                                                                                                                                EPO.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    28, CREATED)
28, LAST SEQUENCE UPDATE)
28, LAST ANNOTATION UPDATE)
160
                                                                                                                                                                                                  188
183
183
55
46
60
                                                                                                                                                                                                                                                                                                                                                                                                                                                            CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
                                                                                                                                                                                         20608
                                                                                                                                              Conservative Substitutions
                                                                                                                                                          Matches
                                                                                                                                                                  Optimized Score
                                                                                                                                                                                                                                                                                     GLYCOPROTEIN;
                                                                                                                                                                                                                                                                                                                                                                                                       TRACY T.E.,
                                                                                                                                                                                         WW;
                                                                                                                                                                                       BY SIMILARITY.
EXYTHROPOITTIN.
BY SIMILARITY.
BY SIMILARITY.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
170
                                                                                                                                                                                                                                                                                                                                                                                                       MULCAHY L.S.,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               188 AA.
                                                                                                                                                         11 11
                                                                                                                                                                                                                                                                                     HORMONE;
180
                                                                                                                                                          533
                                                                                                                                                                                                                                                                                      SIGNAL.
                                                                                                                                                         Significance
Mismatches
 190
                                                                                                                                                                                                                                                                                                                                                                                                        CZELUSNIAK
                                                                                                                                                                                                                                                                                                                                          ADULT MAMMALS
 200
                                                                                                                                                                                                                                                                                                                                                                                                        .
.
                                                                                                                                                 0 0 0
                                                   LGTQLPPQG
                                                                                                                                                9.19
128
0
                                                                                                                                                                                                                                                                                                                                                              SHI HI
```

Listing for Mary Hale

> Tue Apr 11 13:27:37

Page

2

```
210
SA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                     US-08-223-263-1 (1-353)
EPO_CANFA ERYTHROPOIETIN PRECURSOR
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 -KEATSLPEA--TSAAPLRTFTV
140 150
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      RTTAHKDPNAIFLSFOHLLRGKVRFLMLVGGSTLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTA
                                                                                                                                                                                                                                                                                                                                                                                                                      EPO CANFA
P33707;
01-FEB-1994
                                                                                                                                                                 WEN D., BOISSEL J.P.R., TRACY T.E., MULCHY L.S., CZELUSNIAK J., GOODMAN M., BUNN H.F.;
BLOOD 82:1507-1516(1993).
-!- FUNCTION: ERYTHROPOLETIN IS THE PRINCIPAL HORMONE INVOLVED IN REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE PHYSIOLOGICAL LEVEL OF CITCULATING ERYTHROCYTE MASS.
-!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMA AND BY LIVER OF FETAL OR NEONATAL MAMMALS.
                                                                                                                                                                                                                                                                                                                                                                              01-FEB-1994 (REL. 28, CREATED)
01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
01-FEB-1994 (REL. 28, LAST ANNOTATION UPDAT
ERYTHROPOIETIN PRECURSOR (FRAGMENT).
                                                                                                                                             PROSITE;
                                                                                                                                                         EMBL;
                                                                                                                                                                                                                                                                                      SEQUENCE FROM N.A.
93372347
                                                                                                                                                                                                                                                                                                                            EUTHERIA;
                                                                                                                                                                                                                                                                                                                                        CANIS FAMILIARIS (DOG).
EUKARYOTA; METAZOA; CHORDATA;
                                                                                                                               ERYTHROCYTE MATURATION;
                                                                                                                                             L13027; CFERYPRE.
                                                                                                                                                                                                                                                                                                                             CARNIVORA
                                                                                                                                                                                                                                                                                                                                                                                                                                               STANDARD;
                                                                                                                              GLYCOPROTEIN; HORMONE;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 -DTLC--KLFRIYSNFLRGKLTLYTGEACRRGDR
160 170 180 X
                                                                                                                                                                                                                                                                                                                                        VERTEBRATA;
                                                                                                                                                                                                                                                                                                                                                                                                                                               PRT;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                        (FRAGMENT)
                                                                                                                                                                                                                                                                                                                                                                                                                                               175 AA.
                                                                                                                                                                                                                                                                                                                                         TETRAPODA; MAMMALIA;
                                                                                                                               SIGNAL
                                                                                                                                                                                              ADULT MAMMALS
                                                                                                                                                                                                                       OF Z
```



Initial Residue Gaps

Score Identity

11 11 11

24 28% 32

Optimized Score Matches

41 53

Significance Mismatches

8

A D 11

.70 98 0

Conservative Substitutions

5

CARBOHYD CARBOHYD NON TER SEQÜENCE

AA;

19193

MW;

155009

ÇŽ;

POTENTIAL.

DISULFID DISULFID CARBOHYD

<1 23 29 51 46 60 105 175

22 >175 >175 >175 175 175 175

ERYTHROPOIETIN.
BY SIMILARITY.
BY SIMILARITY.
POTENTIAL.
POTENTIAL.

BY SIMILARITY

NON TER

CHAIN

MELTELLLVVMLLLTARLTLSSP---APP--ACDLRVL---SKLLRDSHVLHSRLSQCPEVHPLPTPVLLPA

30

40

50

60

120 -SGQVRL--LLGAL-

Listing 70 Mary Hale

À 3

.

A.

Q

8)

Tue Apr 11 13:27:37 1995

Page 45

US-08-223-263-1 (1-353)
KIKL_MOUSE GLANDULAR KALLIKREIN K22 PRECURSOR (EC 3.4.21.35)

KLKL MOUSE P15948; MUS MUSCULUS (MOUSE).
EUKARYOTA; METAZOA; C
EUTHERIA; RODENTIA. 01-APR-1990 (REL. 14, CREATED)
01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
GLANDULAR KALLIKREIN K22 PRECURSOR (EC 3.4.21.35) (TISSUE (MGK-22) (EPIDEMAL GROWTH FACTOR-BINDING PROTEIN TYPE A)
(NERVE GROWTH FACTOR BETA CHAIN ENDOPEPTIDASE) (BETA-NGF-ENDOPEPTIDASE). STANDARD; CHORDATA; VERTEBRATA; PRT; 259 B TETRAPODA; MAMMALIA; E KALLIKREIN)) (EGF-BP A)

נון SEQUENCE FROM N.A. מיים יות האום/כ; TISSUE=SALIVARY GLAND; DRINKWATER C.C., EVANS B.A., RICHARDS BIOCHEMISTRY 26:6750-6756(1987).

SEQUENCE OF 25-54 LOPEZ G.A.,

FAHNESTOCK M., WOO J.E., MOBLEY W.C.; BIOCHEMISTRY 30:3443-3450(1991). MONS J., WALZ D.A., ARICI M.J.,

SEQUENCE OF 25-41

TISSUE-SURMANDIBULAR GLAND; PETERS J., TAKAHASHI S., TADA M., J. BIOCHEM. 111:643-648(1992). SEQUENCE OF 17-54 AND 70-120 FROM N.A.

MIYAKE

EVANS

EVANS B.A., DRINKWATER C.C., RICHARDS R.I.;

J. BIOL. CHEM. 262:8027-8034 (1987).

-i- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS IN KININOGEN TO RELEASE LYS-BRADYKININ.

-i- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-|-XAA BONDS IN SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF MET-|-XAA OR LEU-|-XAA.

MET-|-XAA OR LEU-|-XAA.

-i- SIMILARITY: TO OTHER TRYPSIN-LIKE SERINE PROTEASES. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.

EMBL; M17977; MMGFBPE1.

EMBL; M17979; MMGFBPE3. ÷

-

- 277

Ž.

Š. *

À y 1/2

Listing for Mary Hale

Tue Apr 11 13:27:37 1995

46

```
Residue Identity
                                                                                                                                                                                                                                                                                                                                                                                                                                        Initial
                                                                                                                                                                                                                                                                                                                                                                                                                                                                         PROSITE; PSO0133; TRYPSIN HIS.
PROSITE; PSO0135; TRYPSIN SER.
HYDROLASE; SERINE PROTEASE; KININGGENASE; GI.
MULTICENE FAMILY; ZYMOGEN; SIGNAL.

SIGNAL 1 1 7 ACTIVATION PEPT CHAIN 25 259 GLANDULAR KALLII ACT SITE 118 118 CHARGE RELAY SY.
ACT SITE 211 211 CHARGE RELAY SY.
ACT SITE 211 211 CHARGE RELAY SY.
ACT SITE 118 118 CHARGE RELAY SY.
ACT SITE 211 211 CHARGE RELAY SY.
ACT SITE 211 211 BY SIMILARITY.
DISULFID 30 171 BY SIMILARITY.
DISULFID 150 217 BY SIMILARITY.
DISULFID 150 217 BY SIMILARITY.
DISULFID 207 232 BY SIMILARITY.
CARBOHYD 102 102 PROBABLE.
SEQUENCE 259 AA; 28384 MW; 358132 CN;
                                                                                                                                                                                        HPDFNMSLLOSVPTGADLSNDLMLLRLSKPADITDVV--KPIDLPTTEPKLGSTCLASGWGSINGLIYQ---
100 110 120 130 140 150 160
                                                                                                                                                                                                                                                                                                      KNSQPWQVAVYYLDEYLCGGVLLDRNWVLTA-
40 X 50 60
                                                                                                    NPNDLQCVSIKLHPNEVCVKAHILKVTDVMLCAGEMNGGKDTCKGDSGGPLIC---
170 180 190 200 210
                                                                                                                                                  130 140 150 160 170 180 190 LIGALQSILIGTQLPPQGRTTAH-KDPNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPPTTAVPSRTSLVLT
                                                                                                                                                                                                                                                   60 100 110 80 90 100 110 HPLPTPVLLPAV--DFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLG-
                    PCGEPN--
                                               200 210 220 230 240 LNELPHRTSGLLETNFTASA-RTTGSGLLKWQQGFRAKIPGLLNQTSRSL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           EMBL; M18618; MMKALF
PIR; A29746; A29746.
PIR; A38356; A38356.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             M18598;
                                                                                                                                                                                                                                                                                                                                   X 10

MELTELLIVVMLL----LTARUTLSSPAPPACD--LRVLSK--LLRDSHVLHSRLSQCPEV
                                                                                                                                                                                                                                                                                                                                                                                                           0 8
APAIYTKLIKFTSWIKDTMAKNP
240 250 X
                                                                                                                                                                                                                                                                                                                                                                                                       23
23
45
                                                                                                                                                                                                                                                                                                                                                                                                       Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                       Optimized Score
Matches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ACTIVATION PEPTIDE.
GIANDULAR KALLIKREIN F
CHARGE RELAY SYSTEM.
CHARGE RELAY SYSTEM.
CHARGE RELAY SYSTEM.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
                                                                                                                                                                                                                                                                                                      -AHCYEDKYNIWLGKNKLFQDEPSAQHRLVSKSFP
70 80 90
                                                                                                                                                                                                                                                                                                                                                                                                                                        (I
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          GLYCOPROTEIN;
                                                                                                                                                                                                                                                                                                                                                                                                                         59
59
                                                                                                                                                                                                                                                                                                                                                                                                                       Significance
Mismatches
                                                                                                   -DGVLQGITSWGST
220 230
                                                                                                                                                                                                                                                                                                                                                                                                         0 0
                                                                                                                                                                                                                                                     -QLSGQVRI
                                                                                                                                                                                                                                                                                                                                                                                                                                       α
                                                                                                                                                                                                                                                                                                                                                                                                       3.22
142
0
```

120

US-08-223-263-1 (1-353) FHUB_ECOLI FERRICHROME TRANSPORT PROTEIN FHUB STANDARD; PRT; 659 A PRECURSOR

BARARA 01-APR-1988 01-OCT-1994 FERRICHROME FHUB ECOLI P06972; 01-APR-1988 (REL. 07, CREATED)
(REL. 07, LAST SEQUENCE UPDATE)
(REL. 30, LAST ANNOTATION UPDATE)
TRANSPORT PROTEIN FHUB PRECURSOR.

Ż

.

ENTEROBACTERIACEAE ESCHERICHIA COLI.
PROKARYOTA; GRACILICUTES;

SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;

Tue Apr 11 13:27:37 1995

47

Page

```
Residue
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Initial
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              SO SO SE LA COMPANIA DE LA COMPANIA DEL COMPANIA DE LA COMPANIA DEL CO
180
                                                                                                                                                                                                                                                                                                                                                                                RDAHGWTWASGALLEDIMPWRWPRIMAALFAGVML---AVAGCIIQRLTGNPMASPEVLGISSGAAFGVVIM 370 380 390 400 410 420 430
                                             LLMMLQ-ASGDPRMAQVLTWISGSTYNATDAQVWRTGIVMVILLAITPLCRRWLTILPL-GGDT---ARAVG
500 510 520 530 540
                                                                                                                               LLGALQSLLGTQLPPQ--
                                                                                                                                                                                                                                                                                                    LPTPVLLPAVDF--
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ECOGENE; EGIO303; FHUB.
ECOGENE; EGIO303; FHUB.
IRON TRANSPORT; TRANSPORT; INNER MEMBRANE; SIGNAL.
SIGNAL
1 20
FERRICHROME-IRON TRAN
CHAIN 21 659
T-> V (IN REF. 3).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           CONFLICT
SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         MOL.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       EMBL;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                +
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               BURKHARDT R.,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        SEQUENCE OF 1-23 FROM N.A. 88038363
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     FUJITA N., MORI H., YURA T., ISHIHAMA A.; NUCLEIC ACIDS RES. 22:1637-1639(1994).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             94261430
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 STRAIN=K12 / W3110;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   SEQUENCE FROM N.A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    KOESTER W., BRAUN V.;
MOL. GEN. GENET. 204:435-442(1986)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   SEQUENCE FROM N.A.
                                                                                                                                                                                                                                              -FLVPGNAFGWLLPAGSLG-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      Identity
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   DICITRATE TRANSPORT SYSTEM.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED. SIMILARITY: TO THE FECC AND FECD PROTEINS OF THE IRON
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       GEN. GENET. 209:49-55(1987).
FUNCTION: THIS IS ONE OF THE PROTEINS INVOLVED IN AFFINITY TRANSPORT OF IRON(III)-FERRICHROME INTO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     X04319; ECFHUB.
; D26562; EC82K.
; X05810; ECFHUACD.
S07318; S07318.
                                                                                                                                                              130
190
                                                                                                                                                                                                                                                                                                                                                                                                                                                             X 10 20 30 40 50 MELTELL----LVVMLLITARLTLSSPAPPACDLRVLSKLLRDS-HVL-HSRLSQCPEVHP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     0 0 0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               BRAUN V.;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           Ä,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  22
24%
77
                                                                                                                                                                                                                                                                                              70 100 110 120
SLGEWKTQMEETKAQDILGAVILLLEGVMAARGQLGPTCLSSLLGQLSGQVRL
200
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        11 L -> V (IN RE 70335 MW; 2233995 CN;
                                                                                                                                     GRTTAHKD-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               Matches = 79
Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Optimized Score
                                                                                                                                                              140
210
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     FERRICHROME-IRON TRANSPORT L -> V (IN REF. 3).
                                                                                                                                                                                                                    -AAVTLLIIMIAAGRGGFSPHRMLLAGMALSTAFTM
460 470 480 490
                                                                                                                                  -PNAIFLSFQHLLRGKVRFLMLVGGSTLCVRRAPP
220
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 H H
                                                                                                                                                              150
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               59
79
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         į
230
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         Significance
Mismatches
                                                                                                                                                              160
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         N
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  HH
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         THE HIGH-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 PROTEIN FHUB
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               E.COLI
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                (III)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        11 11 11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           7
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  7.74
170
0
250
```

Section.

(4

4

Listing for Mary Hale

Tue Apr 11 13:27:37 1995

8

```
US-08-223-263-1 (1-353)
PYC1_YEAST PYRUVATE C
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  SALVGGLLLVFADWCGRMVLFPFQIPAGLLSTFIGAPYFIYLLRKQSR
620 630 640 650 X
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            260 270 280 290 X 300 HELLNGTRGLF-PGPSRRTLGAPDISSGTSDTGSLPP---NLQPGYSPSPTHPPTGQY
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        MALTPTRIAL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  TTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLDQIPGYLNRI
                                                                                                                                 EMBL; J03889; SCPCB.
PIR; A29233; QYBYP.
PROSITE; PS00188; BIOTIN.
PROSITE; PS00866; CPSASE 1
PROSITE; PS00867; CPSASE 2
             ACT SITE
BINDING
SIMILAR
SIMILAR
SIMILAR
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             PYC1 YEAST
P111<del>5</del>4;
                                                                                                                                                                                                                                                                                              MORRIS C.P., LIM F., WALLACE J.C.;
BIOCHEM. BIOPHYS. RES. COMMUN. 145:390-396(1987).
-!- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,
INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY
ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE
CARBOXYL GROUP TO PYRUVATE IN THE SECOND.
-!- CATALYTIC ACTIVITY: ATP + PYRUVATE + HCO(3)(-) = ADP +
ORTHOPHOSPHATE + OXALOACETATE.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 01-UUI-1989 (REL. 11, CREATED)
01-UUI-1989 (REL. 11, LAST SEQUENCE UPDATE)
01-UCI-1993 (REL. 27, LAST ANNOTATION UPDATE)
PYRUVATE CARBOXYLASE 1 (EC 6.4.1.1) (PYRUVIC
                                                                                                                                                                                        -!- PATHWAY: GLUCONEOGENESIS.
-!- SUBUNIT: HOMOTETRAMER.
-!- COFACTOR: BIOTIN, AND ZINC.
-!- SUBCELIDIAR LOCATION: CYTOPLASMIC.
-!- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSFERASES AND CARBAMYL PHOSPHATE SYNTHETASES.
EMBL, J03889; SCECB.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                              LIM F.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       PYC1 OR PYV.
                                                                         ACT
                                                                                       NP_BIND
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOM
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        SEQUENCE FROM N.A., AND PARTIAL SEQUENCE
                                                                                                                      LIGASE;
                                                                                                                                                                                                                                                                                                                                                                                                                                    SEQUENCE OF 1003-1178 FROM N.A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                    \overline{\sim}
                                                                                                                                                                                                                                                                                                                                                                                                                                                                BIOI.
                                                                                                                                                                                                                                                                                                                                                                                                                                                             MORRIS C.P., OCCHIODORO F., V
. CHEM. 263:11493-11497(1988).
                                                                                                                     MULTIFUNCTIONAL
                              182
312
1135
160
350
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            STANDARD;
               187
312
1135
1330
470
1178
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        CARBOXYLASE 1 (EC 6.4.1.1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          570
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       --LILAACLTATATMTIGPL---SFVGLMAPHIARMMGFRRTMPHI--
                                                                                                                      ENZYME;
BY SIMILARITY.

BIOTIN (BY SIMILARITY).

CARBAMOYL PHOSPHATE SYNTHETASES.

WITH OTHER BIOTIN CARBOXYLASES.

WITH OTHER BIOTIN CARRIER PROTEINS

WITH LIPOAMIDE ACETYLTRANSFERASE.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          580
                                                                                      ATP (POTENTIAL).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            PRT;
                                                                                                                   BIOTIN; GLUCONEOGENESIS; ZINC;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        HEMIASCOMYCETES
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            1178 AA.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                               WALLACE J.C.;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          590
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    CARBOXYLASE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         (PYRUVIC
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          600
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        CARBO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ۳
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    (PCB
                AND
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  ۲.
```

Tue Apr 11 13:27:37 1995

Page 49

```
Initial Score
Residue Identity
REAL PROPERTY OF THE PROPERTY 
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          õS
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   PYC2 YEAST STANDARD; PRT; 1180 AA.

p32327;

01-OCT-1993 (REL. 27, CREATED)

01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)

01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)

01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)

PYRUVATE CARBOXYLASE 2 (EC 6.4.1.1) (PYRUVIC CARBO)

PYRUVATE CARBOXYLASE 2 (EC 6.4.1.1) (PYRUVIC CARBO)

PYRUVATE CARBOXYLASE (BAKER'S YEAST).

EURARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         150 X TEPLPPTLPTPVVQLHPLLPDPSAPTPTPTSPLLNTSYTHSQNL--SQEG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    ----LGALQSLLGT-----QLPPQGRTTAHKDPNAIFLSFQHLLRGKV--RFLMLVGGSTLCVKKA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       PNIPFOMILRGANGVAYSSIPDNAIDHFVKOAKDNSVDIFRVFDALNDLEGIKVGVDAVKKAGGVVEATVCF
30 640 650 660 670 680 690 700
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             KLKSNPSVPHLHDAQGNVINVTKSAPPSGWRQVLLEKGPAEFAR--
510 520 530 540
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     60 70 80 90 100 110
TPVLLPAYDFSLGEWKTQMEETKAQDILGAVTLLLEGVMA--ARGQLGPTCLSSLLGQLSGQVR--
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      RIRGVKTNIPFLLTLLTNPVFIEGTYWGT-FIDDTPQLFQMVSSQNRAQKLLHYLADVADNGSSIKGQIGLP
440 X 450 460 470 480 490 500
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                SGDMLQPGKKYNLDYYLEIAEKIVQMG--THILGIKD-MAGTMKPAAAKLLIGSLRAKYPDLPIHVHT--
710 720 730 740 750
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          PSRTSLVLTLNELPNRTSGL---LETNFTASA--
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  -HDSAGTRVASMTACALAGADVVDVAINSMSGL--TSQPSINALLASLEGNIDTGINVEH
770 780 790 800 810 X 820
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              SEQUENCE
                                                                                                                SEQUENCE FROM N.A., AND CHARACTERIZATION
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       X 10 20 30
MELTELLIVVMLLLTARLTLSSPAPPACDLRVLS-----
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        11 (1
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          1178 AA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          140 150 160 170 180 --QLPPQGRTTAHKDPNAIFLSFQHLLRGKV--RFIMLVGGSTLCVRRAPPTTAV
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                22
23%
79
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              130099 MW;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          Matches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           Optimized Score =
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       590
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        210
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              7059028
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   600
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          RTTGSGLLKWQ--QGFRA--KIPGLLNQT-SR
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              Ω,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          68
92
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        220
                                                                                                                                                                                                                                                                                                                    CARBOXYLASE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   610
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            (PYRUVIC CARBO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           -KLLRDSHVLHSRLSQCPEVHPLP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          Significance = Mismatches =
                                                                                                                                                                                                                                                                                                                    2)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             ---QVRQFNGTLLM
550 560
                                                                                                                                                                                                                                                                                                                    (PCB
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         50
```

STUCKA R., DEQUIN S., SALMON J.-M., GANCEDO C.; MOL. GEN. GENET. 229:307-315(1991).

Listing for Mary Hale

Tue Apr 11 13:27:37 1995

Page 50

```
Initial Score
Residue Identity
                                                                                                                                RIRGVKTNIPFILTLINPVFIEGT-YWTTFIDDTPQLFQMVSSQNRAQKLLHYIADIAVNGSSIKGQIGLF
                                                                                                                                                                                                                                                                                                                                ACT C
 60
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           DUBOIS E., EL BAKKOURY M., GIANSDORFF N., MESSENGUY F., PIERARD SCHERENS B., VIERENDEELS F.;
SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
-!- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,
INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY
ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE
CARBOXYL GROUP TO PYRUVATE IN THE SECOND.
                                                                                                                              CONFLICT
CONFLICT
SEQUENCE
                                                                                                                                                                                                                 CONFLICT
CONFLICT
CONFLICT
                                                                                                                                                                                                                                                                                                                                                                                                              EMBL; X59890; SCPYC2G.
EMBL; Z36087; SCYBR218C.
PIR; S46094; S46094.
                                                                                                                                                                                                                                                                                                                                                                                                                                                               -!- PATHWAY: GLUCONEOGENESIS.
-!- SUBUNIT: HOMOTETRAMER.
-!- COFACTOR: BIOTH, AND ZINC.
-!- SUBCELLULAR LOCATION: CYTOPLASMIC.
-!- INDUCTION: BY GLUCOSE.
                                                                                                                                                                                              CONFLICT
                                                                                                                                                                                                                                                                                                                                                                               PROSITE; PS00188; BIOTIN.
PROSITE; PS00866; CPSASE 1.
PROSITE; PS00867; CPSASE 2.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       STRAIN=S288C
                                                                                                                                                                                                                                                                      CONFLICT
                                                                                                                                                                                                                                                                                                                         BINDING
                                                                                                                                                                                                                                                                                                                                                           MULTIGENE FAMILY.
                                                                                                                                                                                                                                                                                                                                                                              PROSITE;
                                                                                                                                                                                                                                                                                                                                                                                                                                            -!- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSFERASES AND CARBAMYL PHOSPHATE SYNTHETASES.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  -!- CATALYTIC ACTIVITY: ATP + PYRUVATE + HCO(3)(-) = ADP +
                                                                                                                                                              CONFLICT
                                                                                                                                                                         CONFLICT
                                                                                                                                                                                    CONFLICT
                                                                                                                                                                                                                                                             CONFLICT
                                                                                                                                                                                                                                                                                            SIMILAR
                                                                                                                                                                                                                                                                                                      SIMILAR
                                                                                                                                                                                                                                                                                                                SIMILAR
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  SEQUENCE FROM N.A.
                                                                                                                                                                                                                                                                                                                                                                   IGASE;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         ORTHOPHOSPHATE + OXALOACETATE.
                     X 450
                                                                                                                                                                                                                                                                                                                                                                    MULTIFUNCTIONAL
                                                   MELTELLLVVMLLLTARLTLSSPAPPACDLRVLS--
 70
                                                                                                                                                    1001
1155
1178
                                                                                                                               1180
1180
                                                                                   23
65
                                                                                                                              AA;
                                                                                                                                                    831
839
1001
1155
1178
                                                                                                                                                                                                                                                                                           1180
                                                                                                                                         1180
 80
                     460
                                                                                             Optimized Score
Matches
                                                                                   Conservative Substitutions
                                                                                                                              130166
                                                                                                                                                                                                                                                                                                                                                                     ENZYME; BIOTIN; GLUCONEOGENESIS;
                                                                                                                                                                                                                                                 WITH
S ->
D ->
N ->
                                                                                                                                                                                           GTA -
90
                                                                                                                                                                                                                                                                                           WITH OTHER BIOTIN CARBOXYLASES.
WITH OTHER BIOTIN CARRIER PROTEINS AND
                                                                                                                                                                                                                                                                                                                         BIOTIN (BY SIMILARITY).
                                                                                                                                                                                                                                                                                                                CARBAMOYL PHOSPHATE SYNTHETASES
                                                                                                                                                                                                                                                                                                                                      BY SIMILARITY.
                                                                                                                                                                                                                                                                                                                                               ATP (POTENTIAL).
                                                                                                                                                                                                                  , ,
                                                                                                                                          ţ
                                                                                                                                                                         ¦
                                                                                                                                                                                    Ļ
                                                                                                                                                                                                                                       Ŷ
                     470
                                                                                                                                                                                                               H LIPOAMIDE ACETYLTRANSFERASE.

C (IN REF. 1).

E (IN REF. 1).

K (IN REF. 1).

K (IN REF. 1).

OF (IN REF. 1).

T (IN REF. 1).
                                                                                                                                          KVIFTR
100
                                                                                                                                                    REF
                                                                                                                                                                                    REF. 1)
                                                                                                                               S
                     480
                                                                                              68
91
                                                   -KLLRDSHVLHSRLSQCPEVHPLP
                                                                                             Significance = Mismatches =
110
                                                                                                                                         ۲.
                     490
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            PIERARD A.,
                                                                                                                                                                                                                                                                                                                                                                     ZINC;
120
                                                                                    11 11
                     500
                                                               50
                                                                                  7.74
237
0
```

4

51

Page

US-08-223-263-1 (1-353) FOR4_MOUSE FORMIN 4. MOUSE

FOR4 MOUSE STANDARD; PRT; 1206 AA. Q05839; Q05839; Q01-JUN-1994 (REL. 29, CREATED) Q1-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE) Q1-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE) FORMIN 4. MUS MUSCULUS (MOUSE). EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; EUTHERIA; RODENTIA. TETRAPODA;

MAMMALIA;

SEQUENCE FROM N.A.

GRUSBY-JACKSON L., KUO A., LEDER P.;
GENES DEV. 6:29-37(1992).
GENES DEV. 6:29-37(1992).
-I- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB
-I- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB
AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR
HE INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.
-I- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT
HAS A FUNCTIONAL ROLE ONLY IN THE LIVER AND LIMB.
-I- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL

<u>-</u> LIMB BUD.

ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING

.

Listing for Mary Hale i

Tue Apr 11 13:27:37 1995

C.

Page

52

Residue Identity Initial SSSEREEES DA PAC US-08-223-263-1 (1-353) FORM MOUSE FORMIN. EMBL; X62377; PRICE CONTROL OF THE C FPLPPTLPTPVV--QLHPLLPDPSAPTPTPTS-RLQAVWPPPKTKDTEEKVGLKYTEAEYQAAI----LHLKREHKEEIETLQAQFELKTFHIRGEHALVTAR--480. 490 500 510 520 530 EQLSQLINIDMPRTEQKEADPEFHGADEMGYSTDQESHKSP-RDAHVQGGQVKARTPETALEAFKALFIRP-350 X 360 370 380 390 400 410 GYLNRIHELLNGTRGLFPGPSRRTLGAP~DISSGTS----DTGSLP--LVLTLNE--FORM MOUSE Q05860; Q1-JUN-1994 Q1-JUN-1994 Q1-JUN-1994 ----LEEAIENLKQQLEKRREGCEEMROVCISTDDDCSEKAFRNVCIQTDRETFLKPCDAESKATRS-SQI-190 EMBL; X62379; MMIFOR. PIR; S24407; S24407. --VPKKLTISLTQLSPSKDSKDIHAPFQTREGTSSSSQQKISPPAPPTPPPPPPPPPPPPPPPPPP---L 610 620 630 640 650 660 -PKKGSTADTSELEALKRKKHEK--ESLRAV----FERSKSRPADSPSDPKSPDQSPTEQDDRTPG 420 430 440 450 460 470 DETERMINED TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE. THIS IS ISOFORM 4 AND IS DIFFERENT IN ITS N-TERMINAL TO THE OTHER MOUSE ISOFORMS SO FAR X 10 20 30 40 50 MELTE-----LLIVVMILLTARUTLSSPAPPACDLRVLSKILRDSHVL--HSRLSQCPE 8 8 8 260 550 -LPNRTSGLLE-TNFTASARTTGS----GLLKWQQGFR---AKIPGLLNQTSRSLDQIP STANDARD; 22 23**%** 95 Optimized Score = Matches = Conservative Substitutions 1468 --PLLNTSYTHSQNLSQEG ALTERNATIVE SPLICING 220 AA ŝ 95 95 Significance = Mismatches = PNLQPGYSPSPTHPPTGQYTI 350 7.74 220 0

(REI. 29, CREATED)
(REI. 29, LAST SEQUENCE UPDATE)
(REI. 29, LAST ANNOTATION UPDATE)

* (

```
Residue
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 Initial
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    CC --- FUNCTION: IS IMPORTANT FOR THE MORPHOGENESIS OF LIMB AND KIDNEY CC AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR MAY HE CC INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.

CC --- SUBCELLULAR LOCATION: NUCLEAR.

CC --- SUBCELLULAR LOCATION: NUCLEAR.

CC --- SUBCELLULAR LOCATION: NUCLEAR.

CC --- TISSUE SPECIFICITY: IT IS PRESENT IN THE ADULT KIDNEY, TESTIS, CLIMB, OVARY, BRAIN, SMALL INTESTINE, SALLVARY GLAND AND HARDERIAN COMPAND. IT IS PRESENT THROUGHOUT THE EMBRYO.

CC GLAND. IT IS PRESENT THROUGHOUT THE EMBRYO.

CC GLAND. IT IS PRESENT THROUGHOUT THE EMBRYO.

CC IS EXPRESSED IN THE APICAL ECTODEWAL RIDGE AND THE MESENCHYMAL COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING KINNEY MORPHOSENESIS, EXPRESSION IS INITIALLY RESTRICTED TO COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.

CC --- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND CONTROL SIZE TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.

R PHRI; $11515; $11515.
                                                                                                                                                                                                                                                                                                                                                                     SADTLEPSSTTKVTETKGASPTSLRASQTWLVSEEASEKGLGPEKITAPPQHQLPPGIASEGFPC-DNFKEQ 570 580 590 600 610 620 630
                                                                                                                               ΝĒ-
                                                                                                                                                                             WPPPKTKDTEEKVGLKYTEAEYQAAI---
710 720
                                                                                                                                                                                                                                                                        TAKDLPNKD--GGVWVPGYRAGPPCPFL----LHEEKEKTSRSEL-
0 650 660 670
                                                                                                                                                                                                                                                                                                                       TPVLLPAVDFSLGEWKTQMEETKAQDILGAVTLLLEGVMAARGQLGPTCLSSLLGQ-LSGQVRLLLGALQ--
                                                                                                                                                                                                                                                                                                                                         6
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  DOMAIN
DOMAIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    SEQUENCE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   DOMAIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  DOMAIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   NUCLEAR
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      WOYCHIK R.P., MAAS R.L.,
NATURE 346:850-853(1990)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           90363291
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     TISSUE=KIDNEY, AND TESTIS;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       SEQUENCE FROM N.A.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       EUTHERIA; RODENTIA.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    MUS MUSCULUS (MOUSE).
EUKARYOTA; METAZOA; C
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   Score
                                                                                                                                                                                                                         -SLLGTQLPPQGRTTAHKDPNAIFLSFQHLLR-GKVRFIMLVGGSTLCVRRAPPTTA-VPSRTSIVLTL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Identity
                                      13
                                                                                                                           PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.
298 301 POLY-SER.
861 864 POLY-SER.
869 970 PRO-RICH.
                                                                                                                                                                                                                                                                                                                                                                                                                    X 10 20 30 40 50 MELTELLLYVMLLITARLT-LSSPAPPACDLRVLSKLLRDSHVL----HSRLSQCPEVHPLP
                                                                                                                                                                                                                                                                                                                                                                                                                                                                          11 11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   ij
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   1468 AA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     861
869
977
                                                                                                                                                                                                                                                                                                                                                                                                                                                                   22
23%
86
                                                                                                                                                                                                                                                                                                                                         80
ŭ,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Conservative Substitutions
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Optimized Score =
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Matches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    163809 MW;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        ZELLER R.,
                                                                                                                                                                          ---LHİKREHKEEIETLQAQFEİKTFHIRGEHALVTAR---
730 740 750 760
                                                                                                                                                                                                                                                                                                                                         90
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                POLY-SER.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    11493196
                                                                                                                                                                                                                                          160
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       VOGT T.F.,
                                                                                                                                                                                                                                                                                                                                         100
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Ω
į
                                                                                                                                                                                                                                                                         ---YLDLNPDQSPTEQDDRTPGRLQAV
680 690 700
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     67
95
                                                                                                                                                                                                                                          170
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          LEDER
                                                                                                                                                                                                                                                                                                                                         110
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Mismatches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   Significance
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ۳.
                                                                                                                                                                                                                                          180
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            TESTIS,
ND HARDERIAN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                        11
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     H H
                                                                                                                                                                                                                                                                                                                                                                                                                                                                      7.74
224
0
                                                                                                                                                                                                                                             190
```

Gaps

Conservative Substitutions

30

40

50

8

×

10

ý.

e .

Listing for Mary Hale 100 Tue Apr 11 13:27:37 1995

Page 52

Residue Initial 250 US-08-223-263-1 (1-353)
EPO_RAT ERYTHROPOIETIN PRECURSOR 320 330 TLPTPVV--QLHPLLPDPSAPTPTPTS-PKKLTISLTQLSPSKDSKDIHAPFQTREGTSSSSQQKISPPAPPTPPPLIPPPLIPPPPLPPG---LGPLPP 840 850 890 EEAIENIKQQLEKRREGCEEMRDVCISTDDDCSPKAFRNVCIQTDRETFLKPCDAESKATRS-SQI--770 780 790 800 810 820 IHELLNGTRGLFPGPSRRTLGAP-DISSGTS-EPO RAT P29676; 01-APR-1993 (01-APR-1993 (01-FEB-1994 (CARBOHYD CARBOHYD BIOCHIM. BIOPHYS. ACTA 1171:99-102(1992).

-!- FUNCTION: ERYTHROPOLETIN IS THE PRINCIPAL HORMONE INVOLVED IN THE PRONCITION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.
-!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMALS.
-NO BY LIVER OF FETAL OR NEONATAL MAMMALS.
-!- SUBCELLULAR LOCATION: SECRETED. DISULFID CARBOHYD EMBL; D10763; RNEPO. PIR; S28148; S28148. PROSITE; PS00817; EPO. SIGNAL NAGAO M., SUGA H., SASAKI R.; EPO. SEQUENCE CHAIN ERYTHROCYTE 9304201 STRAIN=WISTAR; TISSUE=KIDNEY; EUTHERIA; RATTUS NORVEGICUS (RAT). EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; ERYTHROPOIETIN PRECURSOR. SEQUENCE FROM N.A. Identity 260 RODENTIA. E MATURATION;
1 26
27 192
33 187
50 50
64 64
109 109 H # 11 1 27 33 50 64 109 (REL. 25, CREATED)
(REL. 25, LAST SEQUENCE UPDATE)
(REL. 28, LAST ANNOTATION UPDATE) AA; STANDARD; 21 27**%** 24 270 OKANO M., MASUDA S., 21286 Optimized Score Matches 4; GLYCOPROTEIN; HORMONE; S
5 BY SIMILARITY.
5 ERYTHROPOLETIN.
7 BY SIMILARITY.
9 BY SIMILARITY.
9 BY SIMILARITY.
9 BY SIMILARITY.
9 BY SIMILARITY.
1286 MW; 179242 CN; ----DTGSLP-280 -PLLNTSYTHSQNLSQEG 192 NARITA ₽ ----VLALPNSGGPPPPPPPPP -PNLQPGYSPSPTHPPTGQYTLFPLPF 44 53 290 н., SIGNAL. 350 Significance Mismatches IKURA ۳.; 0 0 1.25 118 0 OF / 310

* 1

*

4

Tue Apr 11 13:27:37 1995

10. US-08-223-263-1 (1-353) EPO_MOUSE ERYTHROPOIETIN PRECURSOR.

SO TITITIES DE PRESENTATION DE LA COCCOCCION DE LA COCCOCCION DE LA COCCOCCION DE LA COCCION DE LA C EPO MOUSE P07321; 01-APR-1988 01-APR-1988 01-FEB-1994 01-APR-1988 (REL. 07, CREATED)
01-APR-1998 (REL. 07, LAST SEQUENCE UPDATE)
01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
ERYTHROPOIETIN PRECURSOR. SHOEMAKER C.B., MITSOCK L.D.; MOL. CELL. BIOL. 6:849-858(1986). 87039105 MUS MUSCULUS (MOUSE). EUKARYOTA; METAZOA; C EUTHERIA; RODENTIA. SEQUENCE FROM N.A. STANDARD; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; 192 AA

87039104 SEQUENCE FROM N.A.

MCDONALD J.D., LIN F.-K., GOLDWASSER E.;

MOL. CELL. BIOL. 6:842-848 (1986).

-!- FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN THE REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.

-!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADDIT MAMMALS AND BY LIVER OF FETAL OR NEONATAL MAMMALS.

-!- SUBCELLULAR LOCATION: SECRETED.

EMBL;

L; M12482; MMERP. L; M12930; MMERPA. ; A24901; A24901. ; A24902; A24902. SITE; PS00817; EPO.

ROSITE;

ERYTHROCYTE MATURATION; GLYCOPROTEIN; HORMONE; SIGNAL.

SIGNAL

1 26
CHAIN
27 192
ERYTHROPOIETIN.
DISULFID
33 187
ERYSHROPOIETIN.
CARBOHYD
50 50
BY SIMILARITY.
CARBOHYD
64 64
BY SIMILARITY.
CARBOHYD
109 109
BY SIMILARITY. SEQUENCE ξ 21365 ¥ ERYTHROPOIETIN.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY. 188541 CN;

.

Ý,

13. 36

Page 55

Listing for Mary Ha 6

*.

Tue Apr 11 13:27:37 1995

Page 56

US-08-223-263-1 (1-353)
GAG_MLVAB GAG POLYPROTEIN (CONTAINS: CORE PROTEIN P15; INNER

GAG MLVAB

STANDARD;

PRT;

235

A

21-JUL-1986 (REL. 01, CREATED)
21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
GAG POLYPROTEIN (CONTAINS: CORE PROTEIN P15;

CORE SHELL PROTEIN P30) INNER COAT PROTEIN P12;

GAG.
ABELSON MURINE LEUKEMIA VIRUS.
VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; ONCOVIRINAE.

RETROVIRIDAE;

SEQUENCE FROM N.A. 83221648

REDDY E.P., SMITH M.J., SRINIVASAN A.;
PROC. NATL. ACAD. SCI. U.S.A. 80:3623-3627(1983).
PROC. NATL. ACAD. SCI. U.S.A. 80:3623 IN VIVO YIELD MATURE PROTEINS.
-!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
-!- THIS PROTEIN IS SYNTHESIZED AS A GAG-ABL POLYPROTEIN.

EMBL; VO1541; REAMIV.

PIR; A03931; FOMVGM.

COAT PROTEIN; CORE PROTEIN; POLYPROTEIN; MYRISTYLATION.

COAT PROTEIN; 2

131

CORE PROTEIN p15. CHAIN CHAIN CHAIN LIPID 2 132 216 2 131 215 235 INNER COAT PROTEIN CORE SHELL PROTEIN MYRISTATE. P12

SEQUENCE 235 AA; 25641 MW; 303652 CN;

Gaps Residue Identity Initial Score 170 180 190 200 210 TLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLD 11 11 11 21 23* 29 Conservative Matches Optimized Score Substitutions n 42· Significance Mismatches n 7.25 106 0

* 5.00

当の歌

....

+25

,

Listing for Mary Ha ð

Tue Apr 11 13:27:37

Page 57

MGQTVTTPLSLTL -GHWKDVERIAHNQSVDVKKRRWVTFCSAEWP 20 30 40 -TFNVGW 50

12. US-08-223-263-1 (1-353)
GAG_MSVMT GAG POLYPROTEIN (CONTAINS: CORE PROTEIN P15; INNER

SO FITTER DRACE CONTRACTOR OF THE TENTON OF GAG MSVMT STANDARD; PRT; 468 AA. p32594; CREATED) 01-OCT-1993 (REL. 27, CREATED) 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE) 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE) 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE) CREATED COLYPROTEIN (CONTAINS: CORE PROTEIN P15; MOLONEY MURINE SARCOMA VIRUS (STRAIN TS110). VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; CORE SHELL PROTEIN P30). ONCOVIRINAE. SEQUENCE FROM N.A. 92365121 INNER COAT RETROVIRIDAE; PROTEIN P12;

HUAI L., CHI CHIOCCA S.M., GILBRETH

 J. VIROL. 66:5329-5337(1992).
 -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 -!- THIS PROTEIN IS PROBABLY TRANSLATED AS A GAG-MOS POLYPROTEIN. м.А., AINSWORTH J.R., BISHOP

EMBL;

; M96854; MMSAAX.
; S42321.
A42745; FORVEU.
PROTEIN; CORE PROTEIN; POLYPROTEIN; MYRISTYLATION.
PROTEIN 2 131 CORE PROTEIN P15.
N 132 215 INNER COAT PROTEIN P12.
N 216 468 CORE SHELL PROTEIN P30.

CHAIN CHAIN LIPID SEQUENCE 468 AA; 52681 MW; INNER COAT PROTEIN P12. CORE SHELL PROTEIN P30. MYRISTATE (BY SIMILARIT YRISTATE (BY SIMILARITY). 1077859 CN;

Residue Score Identity H 0 0 21 22**%** 15 Conservative Substitutions Optimized Score Matches li 35 40 Mismatches Significance

170 180 190 200 210 TLCVRRAPPTTAVPSRTSLVLTLNELPNRTSGLLETNFTASARTTGSGLLKWQQGFRAKIPGLLNQTSRSLD MGQTVTTPLSLTL X 10 -DHWKDVERIAHNQSVDVKKRRWVTFCSAEWP 40 -TFNVGW

Listing for Mary Hale

Tue Apr 11 13:27:37 1995

58

13. US-08-223-263-1 (1-353) OCH1_YEAST PROBABLE A

ALPHA-1, 6-MANNOSYLTRANSFERASE ŒC 2.4.1.-

OCHI YEAST P31755; STANDARD; PRT; 480 A

01-JUL-1993 01-JUL-1993 01-FEB-1994

01-JUL-1993 (REL. 26, CREATED)
01-JUL-1993 (REL. 26, LAST EXQUENCE UPDATE)
01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
PROBABLE ALPHA-1, 6-MANNOSYLTRANSFERASE (EC 2.4.1.-).

OCH1.

SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.

SEQUENCE FROM N.A.

STRAIN=EHF-2C;

NAKAYAMA K.-I., NAGASU T., EMBO J. 11:2511-2519(1992) SHIMMA Y.-I., KUROMITSU J.-R., JIGAMI Y.;

-!- FUNCTION: INVOLVED IN OUTER CHAIN ELONGATION OF ASPARAGINE-I OLIGOSACCHARIDES OF THE TYPE MAN(9)GLCNAC(2).
-!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ANCHORED TO ASPARAGINE-LINKED

-!- SUBCELLULAR LOCATION: TYPE II MEMB THE ER AND/OR THE GOLGI MEMBRANES.

EMBL;

EMBL; \$40019; \$40019. D11095; \$COCH1.

PIR; S22701; S22701

TRANSFERASE; GLYCOPROTEIN. GLYCOSYLTRANSFERASE; TRANSMEMBRANE; SIGNAL-ANCHOR,

SO TITIES CARBOHYD CARBOHYD SEQUENCE CARBOHYD CARBOHYD DOMAIN 16 15 30 31 480 3 203 281 341 341 35155 MW; POTENTIAL. POTENTIAL. POTENTIAL. POTENTIAL. CYTOPLASMIC (POTENTIAL).
SIGNAL-ANCHOR (TYPE-II M
LUMENAL (POTENTIAL). (TYPE-II MEMBRANE PROTEIN)

480

1228766 CN;

TRANSMEM DOMAIN

Residue Gaps Initial Identity 0 0 0 Matches = 73 Conservative Substitutions Optimized Score 8 56 73 Significance Mismatches 7.25 236 0

é,

Page 59

```
EG ×
DGSDIMNWTGPG
X 340
                                                                                                                                                                                                                             PIDSWPSQNKSWLNNIIDLNKPIP-YKNSKPSLLSSDEISHQPGLVIGIEADPDRDDWSEWYARRIQFCQWT 200 210 220 230 240 250 260
                                                                                                                                                                                                                                                                                  IQAKPGHPILRELILNITATTL--ASVQ-NPGVPVSEMIDPRFEEDYNVNYRHKRRHDETYKHSELKNNKNV 270 280 290 300 310 320 330
                                                                                                                                                      YSPSPTHPPTGQYTLFPLPPTLPTPVVQLHPLLP---
                                                                                                                                                                                                                                                                                                                                                             SGSYSPDYQYSLISDDSIIPFLENLYAPVPIVIQAFKLMPGNILKADFLRYLLLFARGGIYSDMDT--MLLK
                                                                                                                                                                                                                                                                                                                                                                                                                                                          ---NLKKQI TVNKKKNQ-
70
                                                                                                                                                                                                                                                                                                                                                                                                 160 170 180 220
-KVRFLMLVGGSTLCVRRAP-PTTAVPSRTSLVLTLNELFNRTSGLLETNFTASARTTGSGLLK
                                                                                                                                                                                                                                                                                                                                             140
                                                                                                                                                                                                                                                                                                                                                                                                                                                          LHNLRDQLSFAFPYDSQAPIPQRVWQTWKVG-
80 90 100
                                                                                                                                                                                                                                                                                                                                             150
                                                                                                                                                                                                                                                                                                                                             160
                                                                                                                                                                         -DP-
                                                                                                                                                                    330 340 350
--SAPTPTPTSPLLNTSYTHS---QNLSQ
                                                                                                                                                                                                                                                                                                                                             180
                                                                                                                                                                                                                                                                                                                                                                                                                                                          -ADDKNFPSSFRTYQKTW
110 120
                                                                                                                                                                                                                                                                                                                                               90
```

14. US-08-223-263-1 (1-353) GAG_MSVMO POLYPROTEIN R65 (CONTAINS: CORE PROTEIN P15;

```
FITTER ARM REPORT OF THE FITTER ARM TO SEE THE SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECOND SECO
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               J GAG MSVMO
C P03334;
Z 1-JUL-1986 (REL. 01, CREATED)
JT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANOTATION UPDATE)
DE GAG POLYPROTEIN R65 (CONTAINS: CORE PROTEIN P15; INF
GEORGE SHELL PROTEIN P30; NUCLEOPROTEIN P10).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          PROVIRUS, 82039559
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          MOLONEY MURINE VIRIDAE; SS-RNA
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 REDDY
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      ONCOVIRINAE.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         SEQUENCE FROM N.A
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          SARCOMA VIRUS.
A ENVELOPED VIRUSES;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          POSITIVE-STRAND;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              INNER COAT PROTEIN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    RETROVIRIDAE;
```

REDDY E.P., SMITH M.J., AARONSON S.A.; SCIENCE 214:445-450(1981). CLONE 124 CIRCULAR, SEQUENCE FROM N.A.

VAN BEVEREN C., VAN STRAATEN F., GALLESHAW J.A., VERMA I.M.; CELL 27:97-108(1981).

-!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS. PIR; A03932; FOMVM. A03932; POLYPROTEIN; MYRISTYLATION

COAT PROTEIN; CORE PROTEIN; CHAIN 2 131 CHAIN 132 215 CHAIN 216 478 CHAIN 479 534 ; NUCLEOPROTEIN; POLYPROTEI
CORE PROTEIN P15.
INNER COAT PROTEIN P12.
CORE SHELL PROTEIN P30.
NUCLEOPROTEIN P10.

4 16

> Listing for Mary Hale Tue Apr 11 1 13:27:37 1995

Page 8

Gaps Residue Identity [nitial SOFT LIPID VARIANT SEQUENCE Score 2 519 538 0 11 11 3 21 24% 21 2 519 61209 Conservative Substitutions Optimized Score YW; MYRISTATE. R -> K (IN CLONE; 1333981 CN; II 36 43 124) Significance Mismatches II 7.25 114 0

15. MANB_CALSA BETA-MANNANASE / ENDOGLUCANASE A PRECURSOR (MANNAN

MANB CALSA P22533; 01-AUG-1991 01-JUL-1993 01-JUN-1994 01-AUG-1991 (REL. 19, CREATED)
01-JUL-1993 (REL. 26, LAST ESCUENCE UPDATE)
01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
BETA-MANNANASE / ENDOGLUCANASE A PRECURSOR (MANNAN ENDO-1,4-BETA-MANNANASE) (ENDO-1,4-MANNANASE)
MANNOSIDASE A (EC 3.2.1.78) (BETA-MANNANASE) (ENDO-1,4-MANNANASE)
ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULASE)). MANA. CALDOCELLUM SACCHAROLYTICUM PROKARYOTA; SEQUENCE FROM N.A. NOT YET CLASSIFIED STANDARD; PRT; 1331 A

APPL. GIBBS M.D., SAUL D.J., LU APPL. ENVIRON. MICROBIOL. LUTHI E., BERGQUIST P.L.;)L. 58:3864-3867(1992).

SEQUENCE 91247819 OF 1-346 FROM N.A.

LUETHI E., JASMAT N.B., GRAYLING R.A., LOVE D.R., BERGQUIST P.L.;
APPL. ENVIRON. MICROBIOL. 57:694-700 (1991).

-!- FUNCTION: DEGRADATION OF HEMICELLUICSES, THE SECOND MOST ABUNDANT
POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALITIC DOMAINS WITH
MANNANASE AND ENDOGLICANASE ACTIVITIES.

-!- THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREE CELSIUS.

-!- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC
LINKACES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND
GALACTOGLUCOMANNANS.
-!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.



Initial Score = Residue Identity = Gaps = Initial SQ PARTY TO THE PROPERTY OF TH DOMAIN
ACT_SITE
ACT_SITE
CONFLICT
CONFLICT
SEQUENCE -!- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULASE FAMILY A
(FRANILY 5 OF GLYCOSYL HYDROLASES).
-!- SIMILARITY: THE C-TERMINAL PART BELONGS TO CELLULASE FAMILY J
(FAMILY 44 OF GLYCOSYL HYDROLASES).
EMBL; L01257; CSMANAA1.
EMBL; M36063; CSMANABD.
EMBL; M3745; CSMANABD.
EMBL; M3745; B43745.
PIR; B43745; B43745.
PIR; B43745; B43745.
PIR; B43954; A48954.
PROSITE; PS00659; GLYCOSYL HYDROL F5.
HYDROLASE; GLYCOSIDASE; CELLULOSE DEGRADATION; SIGNAL;
MULTIFUNCTIONAL ENZYME. DOMAIN DOMAIN SIGNAL CHAIN DOMAIN DOMAIN DOMAIN DOMAIN 1331 AA; 21 21% 63 41 1331 325 361 518 564 720 780 1331 162 257 338 Optimized Score = Matches = Conservative Substitutions 146892 MW; BETA-MAINANASE / ENDOGLUCANASE A.
CATALYTIC (MANNANASE ACTIVITY).
PRO/SER/THR-RICH (PT BOX).
SUBSTRATE-BINDING (POTENTIAL).
PRO/SER/THR-RICH (PT BOX).
SUBSTRATE-BINDING (POTENTIAL).
PRO/SER/THR-RICH (PT BOX).
CATALYTIC (ENDOGLUCANASE ACTIVITY).
PRO/SER/THR-RICH (PT BOX). NUCLEOPHILE (BY SIMILARITY).

T -> P (IN REF. 2).

TPTPTPT -> ROHOHRO (IN REF. 2). 9861116 CN; 62 Significance 85 Mismatches 0 0 0

MELTELLLVV-

--MLLLTARLTLSSPAPPACDLRVLSKLLRDSHVLHSRL

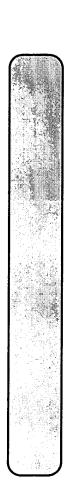
×Į.

Listing for Mary Hale

* Tue Apr 11 13:27:38 1995

8

PTVTPTPTVTATPTPTPTPTPTPTPTISPSPSVVEITINTNAGRT
760 770 780 790 330 340 350 X P-LLPDP----SAPTPTPT-SPLLNTSYTHSQNLSQEG | | | | | | | | | |



maryh@stic

stdin

NeWSprinter20

Tue Apr 11 13:29:56 1995

NeWSprint 2.5 Rev B Openwin library 3 NeWSprint interpreter 210.0

NeWSprint 245